PSYCHOLOGY OF EDUCATION

MAJOR THEMES

Edited by
PETER K. SMITH
and A.D. PELLEGRINI

Also available as a printed book
see title verso for ISBN details

PSYCHOLOGY OF EDUCATION

Major Themes

PSYCHOLOGY OF EDUCATION

Major Themes

VOLUME III

The school curriculum

Edited by

Peter K. Smith and A. D. Pellegrini

London and New York

First published 2000
by RoutledgeFalmer
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by RoutledgeFalmer
29 West 35th Street, New York, NY 10001

This edition published in the Taylor & Francis e-Library, 2004.

RoutledgeFalmer is an imprint of the Taylor & Francis Group

Editorial material and selection © 2000 Peter K. Smith and A. D. Pellegrini;
individual owners retain copyright in their own material.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Psychology of education : major themes / [edited by] Peter K. Smith and A. D. Pellegrini.
p. cm.
Includes bibliographical references and index.
Contents: v. 1. Schools, teachers, and parents — v. 2. Pupils and learning — v. 3. The school curriculum — v. 4. Social behaviour and the school peer group.
ISBN 0-415-19302-8 (set) — ISBN 0-415-19303-6 (v. 1) — ISBN 0-415-19304-4 (v. 2) — ISBN 0-415-19305-2 (v. 3) — ISBN 0-415-19306-0 (v. 4)
1. Educational psychology. I. Smith, Peter K. II. Pellegrini, Anthony D.
LB1051 .P7292774 2000
370.15—dc21 00-034476

ISBN 0-203-45502-9 Master e-book ISBN

ISBN 0-203-76326-2 (Adobe eReader Format)
ISBN 0-415-19302-8 (Set)
ISBN 0-415-19305-2 (Volume III)

The publishers have made every effort to contact authors/copyright holders of works reprinted in *Psychology of Education: Major Themes*. This has not been possible in every case, however, and we would welcome correspondence from those individuals/companies whom we have been unable to trace.

CONTENTS

Volume III
The school curriculum

Introduction to Volume III viii

PART XII
Reading, writing, literacy

61 From utterance to text: the bias of language in speech and writing **3**
D. R. OLSON

62 What no bedtime story means: narrative skills at home and school **32**
S. B. HEATH

63 Schooling for literacy: a review of research on teacher effectiveness and school effectiveness and its implications for contemporary educational policies **61**
D. REYNOLDS

64 Rhyme and alliteration, phoneme detection, and learning to read **81**
P. E. BRYANT, M. MACLEAN, L. L. BRADLEY AND J. CROSSLAND

65 Word recognition: the interface of educational policies and scientific research **101**
M. J. ADAMS AND M. BRUCK

66 Understanding of causal expressions in skilled and less skilled text comprehenders **129**
J. OAKHILL, N. YUILL AND M. L. DONALDSON

67 A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers **141**
R. BROWN, M. PRESSLEY, P. VAN METER AND T. SCHUDER

68 Understanding reading comprehension: current and future contributions of cognitive science **186**
R. F. LORCH JR. AND P. VAN DEN BROEK

PART XIII
Mathematics

69 Developing mathematical knowledge 223
L. B. RESNICK

70 Mathematics in the streets and in schools 239
T. NUNES CARRAHER, D. W. CARRAHER AND A. D. SCHLIEMANN

71 Fostering cognitive growth: a perspective from research on mathematics learning and instruction 251
D. DE CORTE

72 Sociomathematical norms, argumentation, and autonomy in mathematics 271
E. YACKEL AND P. COBB

73 Sex differences in mathematical ability: fact or artifact? 294
C. P. BENBOW AND J. C. STANLEY

PART XIV
Science, social science

74 The acquisition of conceptual knowledge in science by primary school children: group interaction and the understanding of motion down an incline 303
C. HOWE, A. TOLMIE AND C. RODGERS

75 On the complex relation between cognitive developmental research and children's science curricula 326
K. E. METZ

76 Qualitative changes in intuitive biology 342
G. HATANO AND K. INAGAKI

77 Developing understanding while writing essays in history 368
J. F. VOSS AND J. WILEY

78 Generative teaching: an enhancement strategy for the learning of economics in cooperative groups 381
M. KOURILSKY AND M. C. WITTROCK

PART XV
Music, art

79 Research on expert performance and deliberate practice: implications for the education of amateur musicians and music students 399
A. C. LEHMANN AND K. A. ERICSSON

80 How can Chinese children draw so well? 419
E. WINNER

PART XVI
Second language learning

81 Bilingualism and education **441**
K. HAKUTA AND E. E. GARCIA

82 Challenging established views on social issues: the power and limitations of research **455**
W. E. LAMBERT

PART XVII
Computers and media in the classroom

83 Annotation: computers for learning: psychological perspectives **479**
P. LIGHT

84 Hypermedia as an educational technology: a review of the quantitative research literature on learner comprehension, control, and style **496**
A. DILLON AND R. GABBARD

PART XVIII
Cooperative group work, peer tutoring

85 Research on cooperative learning and achievement: what we know, what we need to know **533**
R. SLAVIN

86 Cooperative learning in classrooms: processes and outcomes **562**
N. BENNETT

87 Cooperative learning and peer tutoring: an overview **578**
K. TOPPING

PART XIX
Moral education

88 The cognitive–developmental approach to moral education **597**
L. KOHLBERG

89 Kohlberg's dormant ghosts: the case of education **615**
F. K. OSER

PART XX
Special needs

90 Are we teaching what they need to learn? A critical analysis of the special school curriculum for students with mental retardation **643**
J. W. L. TSE

INTRODUCTION TO VOLUME III

This volume is devoted to studies of learning in particular curriculum areas. It also includes some particular curriculum aids (computers, hypermedia) and methods (cooperative learning, peer tutoring), and the issue of the curriculum for pupils with special needs. The readings chosen are summarised in the series introduction. Here, we mention specific alternative works, books or book chapters, which can usefully supplement or update the readings chosen here.

Reading, writing, literacy: a very useful earlier source of research on reading processes is R. C. Anderson and P. D. Pearson, 'A schema-theoretic view of basic processes in reading', in P. D. Pearson (ed.), *Handbook of reading research* (pp. 225–291), New York: Longman, 1985. Among a range of more recent useful books in the area of reading and literacy are: M. Pressley and P. Afflerbach, *Verbal protocols of reading: The nature of constructively responsive reading*, Hillsdale, NJ: Erlbaum, 1995; A. D. Pellegrini and L. Galda, *The Development of school-based literacy*, London: Routledge, 1999 (which particularly draws out the social context of literacy development); and the chapter by M. J. Adams, R. Treiman and M. Pressley, 'Reading, writing and literacy', in I. E. Sigel and K. E. Renninger (eds), *Handbook of child psychology*, Vol. 4 (pp. 275–356), New York: Wiley, 1998.

Mathematics: for more on mathematics education see P. Cobb (ed.), *Transforming children's mathematics education: International perspectives*, Hillsdale, NJ: Erlbaum, 1990; and T. Nunes and P. Bryant, *Children doing mathematics*, Oxford: Blackwell, 1996.

Science, social science: C. J. Howe, *Conceptual structure in childhood and adolescence: The case of everyday physics*, London: Routledge, 1998, considers a series of experimental studies of children's developing understanding of concepts in physics. An interesting article is S. Vosniadou and W. F. Brewer, 'Mental models of the earth: a study of conceptual change in childhood', *Cognitive Psychology*, 24, 535–585, 1992. For more on acquiring biological knowledge, see G. Hatano and K. Inagaki, 'Cognitive and cultural factors in the acquisition of intuitive biology', in D. R. Olson and N. Torrance (eds), *Handbook of education and human development: New models of learning, teaching and schooling* (pp. 683–708), Oxford: Blackwell, 1996. For more

general perspectives see A. Burgen and K. Härnquist (eds), *Growing up with science: Developing early understanding of science*, Göteburg, Sweden: Academia Europaea, 1997; and D. Kayser and S. Vosniadou, *Modelling changes in understanding: Case studies in physical reasoning*, Amsterdam: Elsevier/Pergamon, 1999. In the social sciences, see M. Carretero and J. F. Voss (eds), *Cognitive and instructional processes in history and the social sciences*, Hillsdale, NJ: Erlbaum, 1994.

Music, art: for general psychological background see D. Hargreaves, *Children and the arts*, and E. Winner, *Invented worlds*, Cambridge, MA, and London: Harvard University Press, 1982. H. Gardner, *Artful scribbles*, London: Jill Norman, 1980, is a classic in the area of children's drawings. Specifically on music education, see J. Sloboda, 'The acquisition of musical performance expertise', in K. A. Ericsson (ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 107–126), Mahwah, NJ: Erlbaum, 1996; and S. H. Kennedy, 'Music in the developmentally appropriate integrated curriculum', in C. D. Hart, D. C. Burts and R. Charlesworth (eds), *Integrated curriculum and developmentally appropriate practice*, New York: State University of New York Press, 1997.

Computers and media in the classroom: For another overview of this topic see C. Crook, *Computers and the collaborative experience of learning*, London: Routledge, 1994. An interesting collection containing contributions on mathematics education and more general topics, as well as computational aids to learning, see J. Bliss, R. Säljö and P. Light (eds), *Learning sites: Social and technological resources for learning*, Amsterdam: Elsevier/Pergamon, 1999.

Cooperative group work, peer tutoring: a classic text on one method of cooperative group work is E. Aronson, *The jigsaw classroom*, Beverly Hills, CA: Sage, 1978. For the Johnsons' work (see also Volume IV), see D. W. Johnson and R. T. Johnson, *Learning together and alone: Cooperative, competitive and individualistic learning* (4th edn), Boston, MA: Allyn & Bacon, 1994, and for more on Slavin's approach see R. E. Slavin, *Cooperative learning: Theory, research and practice* (2nd edn), Boston, MA: Allyn & Bacon, 1995. For an experimental study looking at the effect of cooperative group work on ethnic prejudice and bullying (see also Volume IV), see H. Cowie, P. K. Smith, M. Boulton and R. Laver, *Cooperation in the multi-ethnic classroom*, London: David Fulton, 1994. For collections on peer tutoring and related topics, consult H. Foot, M. J. Morgan and R. J. Shute (eds), *Children helping children*, Chichester: Wiley, 1990; and K. Topping and M. Ehly, *Peer assisted learning*, Mahwah, NJ, and London: Erlbaum, 1998.

Moral education: for more on the classic Kohlberg approach see F. C. Power, L. Higgins and L. Kohlberg, *Lawrence Kohlberg's approach to moral education: A study of three democratic high schools*, New York: Columbia University Press, 1989. Another useful book is P. W. Jackson, R. E. Boostrom and D. T. Hansen, *The moral life of schools*, San Francisco: Jossey-Bass, 1993.

Part XII

READING, WRITING, LITERACY

61

FROM UTTERANCE TO TEXT

The bias of language in speech and writing

D. R. Olson

In this far-ranging essay David Olson attempts to reframe current controversies over several aspects of language, including meaning, comprehension, acquisition, reading, and reasoning. Olson argues that in all these cases the conflicts are rooted in differing assumptions about the relation of meaning to language: whether meaning is extrinsic to language—a relation Olson designates as "utterance"—or intrinsic—a relation he calls "text." On both the individual and cultural levels there has been development, Olson suggests, from language as utterance to language as text. He traces the history and impact of conventionalized, explicit language from the invention of the Greek alphabet through the rise of the British essayist technique. Olson concludes with a discussion of the resulting conception of language and the implications for the linguistic, psychological, and logical issues raised initially.

The faculty of language stands at the center of our conception of mankind; speech makes us human and literacy makes us civilized. It is therefore both interesting and important to consider what, if anything, is distinctive about written language and to consider the consequences of literacy for the bias it may impart both to our culture and to people's psychological processes.

The framework for examining the consequences of literacy has already been laid out. Using cultural and historical evidence, Havelock (1973), Parry (1971), Goody and Watt (1968), Innis (1951), and McLuhan (1964) have argued that the invention of the alphabetic writing system altered the nature of the knowledge which is stored for reuse, the organization of that knowledge, and the cognitive processes of the people who use that written language. Some of the cognitive consequences of schooling and literacy in contemporary societies have been specified through anthropological and cross-cultural psychological research by Cole, Gay, Glick, and Sharp (1971), Scribner and

Source: *Harvard Educational Review*, 1977, 47(3), 257–281.

Cole (1973), Greenfield (1972), Greenfield and Bruner (1969), Goodnow (1976), and others.

However, the more general consequences of the invention of writing systems for the structure of language, the concept of meaning, and the patterns of comprehension and reasoning processes remain largely unknown. The purpose of this paper is to examine the consequences of literacy, particularly those consequences associated with mastery of the "schooled" language of written texts.

In the course of the discussion, I shall repeatedly contrast explicit, written prose statements, which I shall call "texts," with more informal oral-language statements, which I shall call "utterances." Utterances and texts may be contrasted at any one of several levels: the linguistic modes themselves—written language versus oral language; their usual usages—conversation, storytelling, verse, and song for the oral mode versus statements, arguments, and essays for the written mode; their summarizing forms—proverbs and aphorisms for the oral mode versus premises for the written mode; and finally, the cultural traditions built around these modes—an oral tradition versus a literate tradition. My argument will be that there is a transition from utterance to text both culturally and developmentally and that this transition can be described as one of increasing explicitness, with language increasingly able to stand as an unambiguous or autonomous representation of meaning.

This essay (a word I use here in its Old French sense: *essai*—to try) begins by showing that theoretical and empirical debates on various aspects of language—ranging from linguistic theories of meaning to the psychological theories of comprehension, reading, and reasoning—have remained unduly puzzling and polemical primarily because of different assumptions about the locus of meaning. One assumption is that meaning is in the shared intentions of the speaker and the hearer, while the opposite one is that meaning is conventionalized in a sentence itself, that "the meaning is in the text." This essay continues by tracing the assumption that the meaning is in the text from the invention of the alphabetic writing system to the rapid spread of literacy with the invention of printing. The consequences of that assumption, particularly of the attempts to make it true, are examined in terms of the development and exploitation of the "essayist technique." The essay then proceeds to re-examine the linguistic, logical, and psychological issues mentioned at the outset; it demonstrates that the controversies surrounding these issues stem largely from a failure to appreciate the differences between utterances and texts and to understand that the assumptions appropriate for one are not appropriate for the other.

The locus of meaning

The problem at hand is as well raised by a quotation from Martin Luther as by any more contemporary statement: scripture is *sui ipsius interpres*—

scripture is its own interpreter (cited in Gadamer, 1975, p. 154). For Luther, then, the meaning of Scripture depended, not upon the dogmas of the church, but upon a deeper reading of the text. That is, the meaning of the text is in the text *itself*.[1] But is that claim true; is the meaning in the text? As we shall see, the answer offered to that question changed substantially about the time of Luther in regard not only to Scripture but also to philosophical and scientific statements. More important, the answers given to the question lie at the root of several contemporary linguistic and psychological controversies. Let us consider five of these.

In linguistic theory, an important controversy surrounds the status of invariant structures—structures suitable for linguistic, philosophical, and psychological analyses of language. Are these structures to be found in the deep syntactic structure of the sentence itself or in the interaction between the sentence and its user, in what may be called the understanding or interpretation? This argument may be focused in terms of the criterion for judging the well-formedness of a sentence. For Chomsky (1957, 1965) the well-formedness of a sentence—roughly, the judgment that the sentence is a permissible sentence of the language—is determined solely by the base syntactic structure of the sentence. Considerations of comprehensibility and effectiveness, like those of purpose and context, are irrelevant to the judgment. Similarly, the rules for operating upon well-formed base strings are purely formal. For Chomsky the meaning, or semantics, of a sentence is also specified in the base grammatical structure. Each unambiguous or well-formed sentence has one and only one base structure, and this base structure specifies the meaning or semantic structure of that sentence. Hence the meaning of a sentence relies on no private referential or contextual knowledge; nothing is added by the listener. One is justified, therefore, in concluding that, for Chomsky, the meaning is in the sentence per se.[2]

The radical alternative to this view is associated with the general semanticists led by Korzybski (1933), Chase (1954), and Hayakawa (1952). They claim that sentences do not have fixed meanings but depend in every case on the context and purpose for which they were uttered. Chafe (1970) offers a more modest alternative to Chomsky's syntactic bias, asserting that the criterion for the well-formedness of a sentence is determined by the semantic structure: a sentence is well-formed if it is understandable to a listener. This semantic structure is necessarily a part of language users' "knowledge of the world," and language can serve its functions precisely because such knowledge tends to be shared by speakers. Thus comprehension of a sentence involves, to some degree, the use of prior knowledge, contextual cues, and nonlinguistic cues.

In his philosophical discussion of meaning, Grice (1957) makes a distinction that mirrors the difference between the views of Chomsky and Chafe. Grice points out that one may analyze either "sentence meaning" or "speaker's meaning." The sentence per se may mean something other than

what a speaker means by the sentence. For example, the speaker's meaning of "You're standing on my toe" may be "Move your foot." In these terms Chomsky provides a theory of sentence meaning in which the meaning of the sentence is independent of its function or context. Chafe, in contrast, offers a theory of intended meaning that encompasses both the intentions of the speaker and the interpretations the hearer constructs on the bases of the sentence, its perceived context, and its assumed function.

But these theories differ not only in the scope of the problems they attempt to solve. My suggestion is that these linguistic theories specify their central problems differently because of differing implicit assumptions about language; Chomsky's assumption is that language is best represented by written texts; Chafe's is that language is best represented by oral conversational utterances.

Psychological theories of language comprehension reflect these divergent linguistic assumptions. Psycholinguistic models of comprehension such as that of Clark (1974) follow Chomsky in the assumption that one's mental representation of a sentence depends on the recovery of the unique base syntactic structure underlying the sentence. Hence, a sentence is given the same underlying representation regardless of the context or purposes it is ultimately to serve. Similarly, Fodor, Bever, and Garrett (1974) have claimed that the semantic properties of a sentence are determined exclusively and automatically by the specification of the syntactic properties and the lexical items of the sentence. The assumption, once again, is that the meaning, at least the meaning worth psychological study, is in the text.

Conversely, a number of researchers (Anderson & Ortony, 1975; Barclay, 1973; Bransford, Barclay, & Franks, 1972; Bransford & Johnson, 1973; Paris & Carter, 1973) have demonstrated that sentence comprehension depends in large part on the context and on the prior knowledge of the listeners. In one now famous example, the sentence, "The notes were sour because the seams were split," becomes comprehensible only when the listener knows that the topic being discussed is bagpipes. Bransford and Johnson (1973) conclude, "What is understood and remembered about an input depends on the knowledge structures to which it is related" (p. 429).

Differing assumptions as to whether or not the meaning is in the text may also be found in studies of logical reasoning. Logical reasoning is concerned with the formulation and testing of the relations that hold between propositions. Such studies are based on models of formal reasoning in which it is assumed that the rules of inference apply to explicit premises to yield valid inferences. Subjects can be tested on their ability to consistently apply these formal rules to various semantic contents, and development can be charted in terms of the ability to apply the rules consistently to the meaning in the text (Neimark & Slotnick, 1970; Piaget, 1972; Suppes & Feldman, 1971).

Studies have shown, however, that formal propositional logic is a poor model for ordinary reasoning from linguistic propositions. Some researchers

(Taplin & Staudenmayer, 1973) have suggested that logic and reasoning are discontinuous because "the interpreted meaning of a sentence is usually not entirely given by the denotative meaning in the linguistic structure of the sentence" (Staudenmayer, 1975, p. 56); factors such as prior knowledge and contextual presuppositions are also important. Analyzing the protocols of graduate students solving syllogisms, Henle (1962) found that errors resulted more often from an omission of a premise, a modification of a premise, or an importation of new evidence than from a violation of the rules of inference. If logic is considered to be the ability to draw valid conclusions from explicit premises—to operate upon the information in the text—then these students were reasoning somewhat illogically. However, if logic is considered to be the ability to operate on premises as they have been personally interpreted, then these students were completely logical in their operations. The critical issue, again, is whether or not the meaning is assumed to be fully explicit in the text.

Theories of language acquisition also reflect either the assumption that language is autonomous—that the meaning is in the text—or that it is dependent on nonlinguistic knowledge. Assuming that language is autonomous and independent of use or context, Chomsky (1965) and McNeill (1970) have argued that an innate, richly structured language-acquisition device must be postulated to account for the child's remarkable mastery of language. Hypothesized to be innate are structures that define the basic linguistic units (Chomsky, 1972) and the rules for transforming these units. Independent of a particular speaker or hearer, these transformations provide the interpretation given to linguistic forms. For example, at the grammatical level, "John hit Mary" is equivalent to "Mary was hit by John," and at the lexical level, "John" must be animate, human, male, and so on. These conclusions seem plausible, indeed inescapable, as long as it is assumed that language is autonomous and the meanings are in the sentences themselves.

Most recent research on language acquisition has proceeded from the alternative assumption that an utterance is but a fragmentary representation of the intention that lies behind it. Thus the meaning of the utterance comes from shared intentions based upon prior knowledge, the context of the utterance, and habitual patterns of interaction. The contextual dependence of child language was emphasized by de Laguna (1927/1970) and Buhler (1934). De Laguna (1927/1970) claimed, "Just because the terms of the child's language are in themselves so indefinite, it is left to the particular context to determine the specific meaning for each occasion. In order to understand what the baby is saying, you must see what the baby is doing" (pp. 90–91).

Recent studies extend this view. Bloom (1970) has shown, for example, that a young child may use the same surface structure, "Mommy sock," in two quite different contexts to represent quite different deep structures or meanings: in one case, the mother is putting the sock on the child; in the other, the child is picking up the mother's sock. The utterance, therefore, specifies

only part of the meaning, the remainder being specified by the perceived context, accompanying gestures, and the like. Moreover, having established these nonlinguistic meanings, the child can use them as the basis for discovering the structure of language (Brown, 1973; Bruner, 1973; Macnamara, 1972; Nelson, 1974). In other words, linguistic structures are not autonomous but arise out of nonlinguistic structures. There is no need, then, to attribute their origins to innate structures. Language development is primarily a matter of mastering the conventions both for putting more and more of the meaning into the verbal utterance and for reconstructing the intended meaning of the sentence per se. In de Laguna's terms, "The evolution of language is characterized by a progressive freeing of speech from dependence upon the perceived conditions under which it is uttered and heard, and from the behavior that accompanies it. The extreme limit of this freedom is reached in language which is written (or printed) and read" (1927, 1970, p. 107). Thus the predominant view among language-acquisition theorists is that while the meaning initially is not in the language itself, it tends to become so with development.

Finally, theories of reading and learning to read can be seen as expressions of the rival assumptions about the locus of meaning. In one view the meaning is in the text and the student's problem is to find out how to decode that meaning (Carroll & Chall, 1975; Chall, 1967; Gibson & Levin, 1975). In fact, the majority of reading programs are based upon the gradual mastery of subskills such as letter recognition, sound blending, word recognition, and ultimately deciphering meaning. The alternative view is that readers bring the meaning to the text, which merely confirms or disconfirms their expectations (Goodman, 1967; Smith, 1975). Thus if children fail to recognize a particular word or sentence in a context, their expectations generate substitutions that are often semantically appropriate. Again, the basic assumption is that the meaning is—or is not—in the text.

To summarize, the controversial aspects of five issues—the structure of language, the nature of comprehension, the nature of logical reasoning, and the problems of learning to speak and learning to read—can be traced to differing assumptions regarding the autonomy of texts. Further, the distinction between utterances and texts, I suggest, reflects the different assumptions that meaning is or is not in the sentence per se.

The beginnings of a literate technology

Let us consider the origin of the assumption that the meaning is in the text and the implications of that assumption for language use. The assumption regarding the autonomy of texts is relatively recent and the language conforming to it is relatively specialized. Utterance, language that does not conform to this assumption, is best represented by children's early language, oral conversation, memorable oral sayings, and the like. Text, language that

does conform to that assumption, is best represented by formal, written, expository prose statements. My central claim is that the evolution both culturally and developmentally is from utterance to text. While utterance is universal, text appears to have originated with Greek literacy and to have reached a most visible form with the British essayists. My argument, which rests heavily on the seminal works of Havelock (1963), McLuhan (1962), and Goody and Watt (1968), is that the invention of the alphabetic writing system gave to Western culture many of its predominant features including an altered conception of language and an altered conception of rational man. These effects came about, in part, from the creation of explicit, autonomous statements—statements dependent upon an explicit writing system, the alphabet, and an explicit form of argument, the essay. In a word, these effects resulted from putting the meaning into the text.

Meaning in an oral language tradition

Luther's statement, that the meaning of Scripture depended not upon the dogmas of the church, but upon a deeper reading of the text, seems a simple claim. It indicates, however, the profound change that occurred early in the sixteenth century in regard to the presumed autonomy of texts. Prior to the time of Luther, who in this argument represents one turning point in a roughly continuous change in orientation, it was generally assumed that meaning could not be stated explicitly. Statements required interpretation by either scribes or clerics. Luther's claim and the assumption that guided it cut both ways: they were a milestone in the developing awareness that text could explicitly state its meaning—that it did not depend on dogma or interpretive context; more importantly, they also indicated a milestone in the attempt to shape language to more explicitly represent its meanings. This shift in orientation, which I shall elaborate later in terms of the "essayist technique," was one of the high points in the long history of the attempt to make meaning completely explicit. Yet it was, relatively speaking, a mere refinement of the process that had begun with the Greek invention of the alphabet.

Although the Greek alphabet and the growth of Greek literacy may be at the base of Western science and philosophy, it is not to be assumed that preliterate people were primitive in any sense. Modern anthropology has provided many examples of theoretical, mythical, and technological systems of impressive sophistication and appropriateness. It has been established that a complex and extensive literature could exist in the absence of a writing system. In 1928, Milman Parry (1971) demonstrated that the *Iliad* and the *Odyssey*, usually attributed to a literate Homer, were in fact examples of oral composition composed over centuries by preliterate bards for audiences who did not read. In turn, it was recognized that large sections of the Bible possessed a similar oral structure. The books of Moses and the Prophets,

for example, are recorded versions of statements that were shaped through oral methods as part of an oral culture.

To preserve verbal statements in the absence of a writing system, such statements would have to be biased both in form and content towards oral mnemonic devices such as "formalized patterns of speech, recital under ritual conditions, the use of drums and other musical instruments, and the employment of professional remembrances" (Goody & Watt, 1968, p. 31). Language is thus shaped or biased to fit the requirements of oral communication and auditory memory (see, for example, Havelock, 1973, and Frye, 1971). A variety of oral statements such as proverbs, adages, aphorisms, riddles, and verse are distinctive not only in that they preserve important cultural information but also in that they are memorable. They tend, however, *not* to be explicit or to say exactly what they mean; they require context and prior knowledge and wisdom for their interpretation. Solomon, for example, introduced the *Book of Proverbs* by saying: "To understand a proverb and the interpretation; the words of the wise and their dark sayings," (Chapter I:6). Maimonides, the twelfth-century rabbi, pointed out in his *Guide of the Perplexed* that when one interprets parables "according to their external meanings, he too is overtaken by great perplexity!" (1963, p. 6).

The invention of writing did not end the oral tradition. Some aspects of that tradition merely coexist with the more dominant literate traditions. Lord (1960) in his *Singer of Tales* showed that a remnant of such an oral culture persists in Yugoslavia. Even in a predominantly literate culture, aspects of the oral tradition remain. Gray (1973) suggested that Bob Dylan represents the creative end of such an oral tradition in Anglo-American culture; the less creative aspects of that tradition show up in the stock phrases and proverbial sayings that play so large a part in everyday conversational language.

With the introduction of writing, important parts of the oral tradition were written down and preserved in the available literate forms. The important cultural information, the information worth writing down, consisted in large part of statements shaped to fit the requirements of oral memory such as the epics, verse, song, orations, and since readers already knew, through the oral tradition, much of the content, writing served primarily for the storage and retrieval of information that had already been committed to memory, not for the expression of original ideas.

Scripture, at the time of Luther, had just such a status. It consisted in part of statements shaped to the requirements of oral comprehension and oral memory. Scripture had authority, but since the written statements were shorn of their oral contexts, they were assumed to require interpretation. The dogma of the Church, the orally transmitted tradition, had the authority to say what the Scripture meant. In this context Luther's statement can be seen as profoundly radical. Luther claimed that the text supplied sufficient context internally to determine the meaning of the passage; the meaning was in the text. What would have led Luther to make such a radical claim? My

suggestion is that his claim reflected a technological change—the invention of printing—one in a series of developments in the increasing explicitness of language, which we shall now examine.

Alphabetic writing—making meanings explicit

Significant oral-language statements, to be memorable, must be cast into some oral, poetic form. Consequently, as we have seen, these statements do not directly say what they mean. With the invention of writing, the limitations of oral memory became less critical. The written statement, constituting a more or less permanent artifact, no longer depended on its "poetized" form for its preservation.

However, whether or not a writing system can preserve the meanings of statements depends upon the characteristics of the system. An elliptical or nonexplicit writing system, like nonexplicit statements, tends to rely on prior knowledge and expectancies. An explicit writing system unambiguously represents meanings—the meaning is in the text. It has a minimum of homophones (seen/scene) and homographs (lead/lead) at the phonemic and graphemic levels, few ambiguities at the grammatical level, and few permissible interpretations at the semantic level.

The Greek alphabet was the first to approach such a degree of explicitness and yet to be simple enough to provide a base for mass literacy. Gelb (1952) differentiated four main stages in the development of writing systems. The first stage, which goes back to prehistory, involves the expression of ideas through pictures and pictographic writing. Such writing systems have been called ideographic in that they represent and communicate ideas directly without appeal to the structure of spoken language. While the signs are easily learned and recognized, there are problems associated with their use: any full system requires some four or five thousand characters for ordinary usage; their concreteness makes the representation of abstract terms difficult; they are difficult to arrange so as to produce statements (Gombrich, 1974); and they tend to limit the number of things that can be expressed.

The next stage was the invention of the principle of phonetization, the attempt to make writing reflect the sound structure of speech. In an attempt to capture the properties of speech, early phonetic systems—Sumerian, Egyptian, Hittite, and Chinese—all contained signs of three different types: word signs or logogens, syllabic signs, and auxiliary signs.

The third stage was the development of syllabaries which did away both with word signs and with signs representing sounds having more than one consonant. Whereas earlier syllabaries had separate signs for such syllables as *ta* and *tam*, the West Semitic syllabaries reduced the syllable to a single consonant-vowel sequence, thereby reducing the number of signs. However, since these Semitic syllabaries did not have explicit representations for vowels, the script frequently resulted in ambiguities in pronunciation, particularly in

cases of writing proper names and other words which could not be retrieved from context. Semitic writing systems thus introduced phonetic indicators called Matres Lectionis (literally: "mothers of reading") to differentiate the vowel sounds (Gelb, 1952, p. 166).

The final stage in the invention of the alphabet, a step taken only by the Greeks, was the invention of a phonemic alphabet (Gelb, 1952; Goody & Watt, 1963). The Greeks did so, Gelb suggests, by using consistently the Matres Lectionis which the Semites had used sporadically. They discovered that these indicators were not syllables but rather vowels. Consequently the sign that preceded the indicator also must not be a syllable but rather a consonant. Havelock (1973) comments: "At a stroke, by this analysis, the Greeks provided a table of elements of linguistic sound not only manageable because of its economy, but for the first time in the history of *homo sapiens*, also accurate" (p. 11).

The faithful transcription of the sound patterns of speech by a fully developed alphabet has freed writing from some of the ambiguities of oral language. Many sentences that are ambiguous when spoken are unambiguous when written—for example, "il vient toujours a sept heures" ("he always comes at seven o'clock") versus "il vient toujours a cette heure" ("he always comes at this hour") (Lyons, 1969, p. 41). However, a fully developed alphabet does not exhaust the possibilities for explicitness of a writing system. According to Bloomfield (1939) and Kneale and Kneale (1962), the remaining lack of explicitness necessitated the invention of the formal languages of logic and mathematics.

To summarize, we have considered the extent to which meaning is explicitly represented in a statement. Oral language statements must be poetized to be remembered, but in the process they lose some of their explicitness; they require interpretation by a wise man, scribe, or cleric. Written statements bypass the limitations of memory, but the extent to which a writing system can explicitly represent meaning depends upon the nature of the system. Systems such as syllabaries that represent several meanings with the same visual sign are somewhat ambiguous or nonexplicit. As a consequence, they again require interpretation by some authority. Statements can become relatively free from judgment or interpretation only with a highly explicit writing system such as the alphabet. The Greek alphabet, through its ability to record exactly what is said, provided a tool for the formulation and criticism of explicit meanings and was therefore critical to the evolution of Greek literacy and Greek culture.

Written text as an exploratory device

Writing systems with a relatively lower degree of explicitness, such as the syllabaries, tended to serve a somewhat limited purpose, primarily that of providing an aid to memory. Havelock (1973) states:

> When it came to transcribing discursive speech, difficulties of interpretation would discourage the practice of using the script for novel or freely-invented discourse. The practice that would be encouraged would be to use the system as a reminder of something already familiar, so that recollection of its familiarity would aid the reader in getting the right interpretation. . . . It would in short tend to be something—tale, proverb, parable, fable and the like—which already existed in oral form and had been composed according to oral rules. The syllabic system in short provided techniques for recall of what was already familiar, not instruments for formulating novel statements which could further the exploration of new experience. (p. 238)

The alphabet had no such limits of interpretation. The decrease in ambiguity of symbols—for example, the decrease in the number of homographs—would permit a reader to assign the appropriate interpretation to a written statement even without highly tuned expectations as to what the text was likely to say. The decreased reliance upon prior knowledge or expectancies was therefore a significant step towards making meaning explicit in the conventionalized linguistic system. The technology was sufficiently explicit to permit one to analyze the sentence meaning apart from the speaker's meaning. Simultaneously, written language became an instrument for the formulation and preservation of original statements that could violate readers' expectancies and commonsense knowledge. Written language had come free from its base in the mother tongue; it had begun the transformation from utterance to text.

The availability of an explicit writing system, however, does not assure that the statements recorded in that language will be semantically explicit. As previously mentioned, the first statements written down tended to be those that had already been shaped to the requirements of oral production and oral memory, the Greek epics being a case in point. Over time, however, the Greeks came to fully exploit the powers of their alphabetic writing system. In fact, Havelock (1973) has argued that the Greeks' use of this invention was responsible for the development of the intellectual qualities found in classical Greece:

> And so, as the fifth century passes into the fourth, the full effect upon Greece of the alphabetic revolution begins to assert itself. The governing word ceases to be a vibration heard by the ear and nourished in the memory. It becomes a visible artifact. Storage of information for reuse, as a formula designed to explain the dynamics of western culture, ceases to be a metaphor. The documented statement persisting through time unchanged is to release the human brain from certain formidable burdens of memorization while increasing

> the energies available for conceptual thought. The results as they are to be observed in the intellectual history of Greece and Europe were profound. (p. 60)

Some of the effects of the Greeks' utilization of the alphabetic writing system are worth reviewing. First, as Goody and Watt (1968) and a number of other scholars have shown, it permitted a differentiation of myth and history with a new regard for literal truth. When the Homeric epics were written down, they could be subjected to critical analysis and their inconsistencies became apparent. Indeed, Hecataeus, faced with writing a history of Greece, said: "What I write is the account I believe to be true. For the stories the Greeks tell are many and in my opinion ridiculous" (cited in Goody & Watt, 1968, p. 45). Second, the use of the alphabetic system altered the relative regard for poetry and for prose. Prose statements were neither subtle nor devious; they tended to mean what they said. Havelock (1963) has demonstrated that Plato's *Republic* diverged from the tradition of the oral Homeric poets and represented a growing reliance on prose statements.

Third, the emphasis on written prose, as in Aristotle's *Analytics* (see Goody & Watt, 1968, pp. 52–54), permitted the abstraction of logical procedures that could serve as the rules for thinking. Syllogisms could operate on prose premises but not on oral statements such as proverbs. Further, the use of written prose led to the development of abstract categories, the genus/species taxonomies so important not only to Greek science but also to the formation and division of various subject-matter areas. Much of Greek thought was concerned with satisfactorily explaining the meaning of terms. And formulating a definition is essentially a literate enterprise outside of the context of ongoing speech—an attempt to provide the explicit meaning of a word in terms of the other words in the system (see, for example, Bruner & Olson, in press; Goody & Watt, 1968; and Havelock, 1976).

The Greeks, thinking that they had discovered a method for determining objective truth, were in fact doing little more than detecting the properties implicit in their native tongue. Their rules for mind were not rules for thinking but rather rules for using language consistently; the abstract properties of their category system were not true or unbiased descriptions of reality but rather invariants in the structure of their language. Writing became an instrument for making explicit the knowledge that was already implicit in their habits of speech and, in the process, tidying up and ordering that knowledge. This important but clearly biased effort was the first dramatic impact of writing on knowledge.

The Greeks' concern with literacy was not without critics. Written statements could not be interrogated if a misunderstanding occurred, and they could not be altered to suit the requirements of listeners. Thus Socrates concluded in *Phaedrus*: "Anyone who leaves behind him a written manual,

and likewise anyone who takes it over from him, on the supposition that such writing will provide something reliable and permanent, must be exceedingly simple minded" (*Phaedrus*, 277c, cited in Goody & Watt, 1968, p. 51). In the *Seventh Letter*, Plato says: "No intelligent man will ever be so bold as to put into language those things which his reason has contemplated, especially not into a form that is unalterable—which must be the case with what is expressed in written symbols" (*Seventh Letter*, 341 c-d, cited in Bluck, 1949, p. 176).

The essayist technique

Although the Greeks exploited the resources of written language, the invention of printing allowed an expanded and heterogeneous reading public to use those resources in a much more systematic way. The invention of printing prompted an intellectual revolution of similar magnitude to that of the Greek period (see McLuhan, 1962, and Ong, 1971, for fascinating accounts). However, the rise of print literacy did not merely preserve the analytic uses of writing developed by the Greeks; it involved as well, I suggest, further evolution in the explicitness of writing at the semantic level. That is, the increased explicitness of language was not so much a result of minimizing the ambiguity of words at the graphemic level but rather a result of minimizing the possible interpretations of statements. A sentence was written to have only one meaning. In addition, there was a further test of the adequacy of a statement's representation of presumed intention: the ability of that statement to stand up to analysis of its implications. To illustrate, if one assumes that statement X is true, then the implication Y should also be true. However, suppose that on further reflection Y is found to be indefensible. Then presumably statement X was not intended in the first place and would have to be revised.

This approach to texts as autonomous representations of meaning was reflected in the way texts were both read and written. A reader's task was to determine exactly what each sentence was asserting and to determine the presuppositions and implications of that statement. If one could assume that an author had actually intended what was written and that the statements were true, then the statements would stand up under scrutiny. Luther made just this assumption about Scripture early in the sixteenth century, shortly after the invention and wide utilization of printing. One of the more dramatic misapplications of the same assumption was Bishop Usher's inference from biblical genealogies that the world was created in 4004 B.C.

The more fundamental effect of this approach to text was on the writer, whose task now was to create autonomous text—to write in such a manner that the sentence was an adequate, explicit representation of the meaning, relying on no implicit premises or personal interpretations. Moreover, the sentence had to withstand analysis of its presuppositions and implications.

This fostered the use of prose as a form of extended statements from which a series of necessary implications could be drawn.

The British essayists were among the first to exploit writing for the purpose of formulating original theoretical knowledge. John Locke's *An Essay Concerning Human Understanding* (1690/1961) well represents the intellectual bias that originated at that time and, to a large extent, characterizes our present use of language. Knowledge was taken to be the product of an extended logical essay—the output of the repeated application in a single coherent text of the technique of examining an assertion to determine all of its implications. It is interesting to note that when Locke began his criticism of human understanding he thought that he could write it on a sheet of paper in an evening. By the time he had exhausted the possibilities of both the subject and the new technology, the essay had taken twenty years and two volumes.

Locke's essayist technique differed notably from the predominant writing style of the time. Ellul (1964) says, "An uninitiated reader who opens a scientific treatise on law, economy, medicine or history published between the sixteenth and eighteenth centuries is struck most forcibly by the complete absence of logical order" (p. 39); and he notes, "It was more a question of personal exchange than of taking an objective position" (p. 41). In the "Introduction" to *Some Thoughts Concerning Education* (Locke, 1880), Quick reports that Locke himself made similar criticisms of the essays of Montaigne. For Locke and others writing as he did, the essay came to serve as an exploratory device for examining problems and in the course of that examination producing new knowledge. The essay could serve these functions, at least for the purposes of science and philosophy, only by adopting the language of explicit, written, logically connected prose.

This specialized form of language was adopted by the Royal Society of London which, according to its historian Sprat (1667/1966), was concerned "with the advancement of science and with the improvement of the English language as a medium of prose" (p. 56). The society demanded a mathematical plainness of language and rejected all amplifications, digressions, and swellings of style. This use of language made writing a powerful intellectual tool, I have suggested, by rendering the logical implications of statements more detectable and by altering the statements themselves to make their implications both clear and true.

The process of formulating statements, deriving their implications, testing the truth of those implications, and using the results to revise or generalize from the original statement characterized not only empiricist philosophy but also the development of deductive empirical science. The result was the same, namely the formulation of a small set of connected statements of great generality that may occur as topic sentences of paragraphs or as premises of extended scientific or philosophical treatise. Such statements were notable not only in their novelty and abstractness but also in that they related to

prior knowledge in an entirely new way. No longer did general premises necessarily rest on the data of common experience, that is on commonsense intuition. Rather, as Bertrand Russell (1940) claimed for mathematics, a premise is believed because true implications follow from it, not because it is intuitively plausible. In fact, it is just this mode of using language—the deduction of counterintuitive models of reality—which distinguishes modern from ancient science (see Ong, 1958).

Moreover, not only did the language change, the picture of reality sustained by language changed as well; language and reality were reordered. Inhelder and Piaget (1958) describe this altered relationship between language and reality as a stage of mental development:

> The most distinctive property of formal thought is this reversal of direction between reality and possibility; instead of deriving a rudimentary theory from the empirical data as is done in concrete inferences, formal thought begins with a theoretical synthesis implying that certain relations are necessary and thus proceeds in the opposite direction. (p. 251)

The ability to make this "theoretical synthesis," I suggest, is tied to the analysis of the implications of the explicit theoretical statements permitted by writing.

Others have made the same point. Ricoeur (1973) has argued that language is not simply a reflection of reality but rather a means of investigating and enlarging reality. Hence, the text does not merely reflect readers' expectations; instead the explicitness of text gives them a basis for constructing a meaning and then evaluating their own experiences in terms of it. Thus text *can* serve to realign language and reality. N. Goodman (1968), too, claims that "the world is as many ways as it can be truly described" (p. 6).

This property of language, according to Popper (1972), opens up the possibility of "objective knowledge." Popper claims that the acquisition of theoretical knowledge proceeds by offering an explicit theory (a statement), deriving and testing implications of the theory, and revising it in such a way that its implications are both productive and defensible. The result is a picture of the world derived from the repeated application of a particular literary technique: "science is a branch of literature" (Popper, 1972, p. 185).

Thus far I have summarized two of the major stages or steps in the creation of explicit, autonomous meanings. The first step toward making language explicit was at the graphemic level with the invention of an alphabetic writing system. Because it had a distinctive sign for each of the represented sounds and thereby reduced the ambiguity of the signs, an alphabetic system relied much less on readers' prior knowledge and expectancies than other writing systems. This explicitness permitted the preservation of meaning across space and time and the recovery of meaning by the more or less

uninitiated. Even original ideas could be formulated in language and recovered by readers without recourse to some intermediary sage.

The second step involved the further development of explicitness at the semantic level by allowing a given sentence to have only one interpretation. Proverbial and poetic statements, for example, were not permissible because they admitted more than one interpretation, the appropriate one determined by the context of utterance. The attempt was to construct sentences for which the meaning was dictated by the lexical and syntactic features of the sentence itself. To this end, the meaning of terms had to be conventionalized by means of definitions, and the rules of implication had to be articulated and systematically applied.

The Greeks perfected the alphabetic system and began developing the writing style that, encouraged by the invention of printing and the form of extended texts it permitted, culminated in the essayist technique. The result was not an ordinary language, not a mother tongue, but rather a form of language specialized to serve the requirements of autonomous, written, formalized text. Indeed, children are progressively inducted into the use of this language during the school years. Thus formal schooling, in the process of teaching children to deal with prose texts, fosters the ability to "speak a written language" (Greenfield, 1972, p. 169).

The effects of considerations of literacy on issues of language

Let us return to the linguistic and psychological issues with which we began and reconsider them in the light of the cultural inventions that have served to make language explicit, to put the meaning into the text.

Linguistic theory

The differences between oral language and written text may help to explain the current controversy between the syntactic approach represented by Chomsky and the semantic approach represented by Chafe. Several aspects of Chomsky's theory of grammar require attention in this regard. For Chomsky, the meaning of language is not tied to the speaker's knowledge of the world but is determined by the sentence or text itself. The meaning of a sentence is assigned formally or mechanically on the basis of the syntactic and lexical properties of the sentence per se and not on the basis of the expectancies or preferred interpretations of the listener (Chomsky, 1972, p. 24). Chomsky's theory is fundamentally designed to preserve the truth conditions of the sentence, and permissible transformations are ones that preserve truth. To illustrate, an active sentence can be related to a passive sentence by means of a set of transformations because they are assumed to share a common base or underlying structure. The equivalence between

active and passive sentences is logical meaning: one sentence is true if and only if the other is true (see Harman, 1972; Lakoff, 1972).

My conjecture is that Chomsky's theory applies to a particular specialization of language, namely, the explicit written prose that serves as the primary tool of science and philosophy. It can serve as a theory of speech only when the sentence meaning is a fully adequate representation of the speaker's meaning. In ordinary conversational language, this is rarely the case. The empirical studies mentioned earlier have provided strong evidence that experimental subjects rarely confine their interpretations to the information conventionalized in text. Rather, they treat a sentence as a cue to a more elaborate meaning.

As we have seen, other linguistic theories treat language as a means of representing and recovering the intentions of the speaker. The general semanticists and, to a lesser extent, Chafe have argued that the linguistic system is not autonomous. The meaning of a sentence is not determined exclusively by the lexical and syntactic properties of the sentence itself; rather, the sentence is an indication of the speaker's meaning. While this assumption seems appropriate to the vast range of ordinary oral language, it overlooks the case in which the intended meaning is exactly represented by the sentence meaning as is ideally the case in explicit essayist prose.

We may conclude, then, that the controversy between the syntacticists and the semanticists is reducible to the alternative assumptions that language is appropriately represented in terms of sentence meanings or in terms of speaker's meanings. The latter assumption is entirely appropriate, I suggest, for the description of the ordinary oral conversational language, for what I have called utterances. On the other hand, I propose that Chomsky's theory is not a theory of language generally but a theory of a particular specialized form of language assumed by Luther, exploited by the British essayists, and formalized by the logical positivists. It is a model for the structure of autonomous written prose, for what I have called text.

On comprehension

The comprehension of sentences involves several different processes. Ordinary conversational speech, especially children's speech, relies for its comprehension on a wide range of information beyond that explicitly marked in the language. To permit communication at all, there must be wide agreement among users of a language as to phonological, syntactic, and semantic conventions. A small set of language forms, however, maps onto an exceedingly wide range of referential events; hence, ambiguity is always possible if not inevitable. Speakers in face-to-face situations circumvent this ambiguity by means of such prosodic and paralinguistic cues as gestures, intonation, stress, quizzical looks, and restatement. Sentences in conversational contexts, then, are interpreted in terms of the following: agreed-upon lexical

and syntactic conventions; a shared knowledge of events and a preferred way of interpreting them; a shared perceptual context; and agreed-upon prosodic features and paralinguistic conventions.

Written languages can have no recourse to shared context, prosodic features, or paralinguistic conventions since the preserved sentences have to be understood in contexts other than those in which they were written. The comprehension of such texts requires agreed-upon linguistic conventions, a shared knowledge of the world, and a preferred way of interpreting events. But Luther denied the dependence of text on a presupposed, commonsensical knowledge of the world, and I have tried to show that the linguistic style of the essayist has systematically attempted to minimize if not eliminate this dependence. This attempt has proceeded by assigning the information carried implicitly by nonlinguistic means into an enlarged set of explicit linguistic conventions. In this way written textual language can be richer and more explicit than its oral language counterpart. Within this genre of literature, if unconventionalized or nonlinguistic knowledge is permitted to intrude, we charge the writer with reasoning via unspecified inferences and assumptions or the reader with misreading the text.

Comprehension, therefore, may be represented by a set of procedures that involves selectively applying one's personal experiences or knowledge of the world to the surface structure of sentences to yield a meaning. In so doing, one elaborates, assimilates, or perhaps "imagines" the sentence. And these elaborative procedures are perfectly appropriate to the comprehension of ordinary conversational utterances. In turn, the sentence becomes more comprehensible and dramatically more memorable, as Anderson and Ortony (1975), Bransford and Johnson (1973), and Bransford, Barclay, and Franks (1972) have shown.

The price to be paid for such elaboration and assimilation is that the listener's or reader's meaning deviates to some degree from the meaning actually represented in the sentence. Such interpretation may alter the truth conditions specified by the statement. To illustrate, using Anderson and Ortony's sentence, if the statement "the apples are in the container" is interpreted as "the apples are in the basket," the interpretation specifies a different set of truth conditions than did the original statement. We could legitimately say that the statement had been misinterpreted. Yet that is what normally occurs in the process of understanding and remembering sentences; moreover, as we have shown in our laboratory, it is what preschool children regularly do (Olson & Nickerson, in press; Pike & Olson, in press; Hildyard & Olson, Note 1). If young children are given the statements, "John hit Mary" or "John has more than Mary," unlike adults, they are incapable of determining the direct logical implications that "Mary was hit by John" or "Mary has less than John." If the sentence is given out of context, they may inquire, "Who is Mary?" Given an appropriate story or pictorial context, children can assimilate the first statement to that context and then give a

new description of what they now know. If the sentence cannot be assimilated to their knowledge base, they are helpless to arrive at its implications; children are unable to apply interpretive procedures to the sentence meaning, the meaning in the text. They can, however, use sentences as a cue to speaker's meaning if these sentences occur in an appropriate context. Literate adults are quite capable of treating sentences in either way. What they do presumably depends on whether the sentence is too long to be remembered verbatim, whether it is written and remains available for repeated consultation, or, perhaps, whether the sentence is regarded as utterance or text.

On reasoning

Extending the argument to reasoning tasks, it is clear that solutions may be reached in either of two quite different ways. Relying on the processes usually involved in the comprehension of spoken language, one may interpret a premise in terms of previous knowledge of the world, operate on that resulting knowledge, and produce an answer other than that expected on a purely formal logical basis. Such reasoning, based on an intrusion of unspecified knowledge, is not a logical argument but an enthymeme. Nevertheless, it is the most common form of ordinary reasoning (Cole, Gay, Glick, & Sharp, 1971; Wason & Johnson-Laird, 1972).

Logical reasoning, on the other hand, is the procedure of using conventionalized rules of language to draw necessary implications from statements treated as text. For such reasoning, the implications may run counter to expectancies or may be demonstrably false in their extension; however, it matters only that the conclusion follows directly from the sentence meaning, the conventionalized aspects of the statement itself. The fact that most people have difficulty with such operations indicates simply their inability or lack of experience in suspending prior knowledge and expectancies in order to honor the sentence meaning of statements. In fact, Henle (1962) has noted that in reasoning tasks subjects often have difficulty in distinguishing between a conclusion that is logically true, one that is factually true, and one with which they agree. According to the analysis offered here, in the first case the conclusion logically follows from the text—the meaning is restricted to that explicitly represented or conventionalized in the text and to the implications that necessarily follow; in the second case the conclusion follows from unstated but shared knowledge of the world; in the third case the conclusion follows from unspecified and unshared personal knowledge. I would argue that in neither of the latter cases are we justified in calling the reasoning logical.

Logical reasoning as defined here assumes that fully explicit, unambiguous statements can be created to serve as premises. This is a goal that consistently evades ordinary language use. It is extremely difficult if not impossible to create statements that specify all and only the necessary and sufficient

information for drawing logical inferences.[3] Hence, formal reasoning has led to a reliance, where possible, on the use of symbols related by a logical calculus. To illustrate the difficulties, I will use three studies from our laboratory. Bracewell (Note 2) has shown that the simple propositional statement employed by Wason and Johnson-Laird (1970), "If p is on one side, then q is on the other," is ambiguous in at least two ways: "one side" may be interpreted as referring to "the showing side" or to "either the showing side or the hidden side"; "if . . . then" may be interpreted as a conditional relation or as a biconditional relation. Differences in subjects' performance can be traced to different interpretations of the proposition. In a similar vein, Hidi (Note 3) has shown that if a simple proposition such as "if you go to Ottawa, you must travel by car" is understood as describing a temporal event, subjects draw quite different inferences than if it is treated purely as a logical statement. In a developmental study, Ford (1976) has shown that, given a disjunctive statement, children (and adults in natural language contexts) treat "or" as posing a simple choice between mutually exclusive, disjoint alternatives (for example, "Do you want an apple *or* an orange?" "An apple."). When children of five or six years of age are presented with "or" commands involving disjoint events as well as overlapping and inclusive events—the latter being involved in Piaget's famous task "Are there more rabbits *or* animals?"—Ford found that children's logical competence breaks down only when the known structure of events runs counter to the presuppositions of the language. Rather than revise their conception of events—rabbits and animals are not disjoint classes—children misinterpret or reject the sentence. They say, for example, "There are more rabbits because there are only two ducks!"

There are, then, at least two aspects to the study of logical reasoning. The first stems from the fact that statements are often ambiguous, especially when they occur out of context. Thus failures in reasoning may reflect merely the assignment of an interpretation that, although it is consistent with the sentence meaning explicit in the text, is different from the one intended by the experimenter. Second, logical development in a literate culture involves learning to apply logical operations to the sentence meaning rather than to the assimilated or interpreted or assumed speaker's meaning. Development consists of learning to confine interpretation to the meaning explicitly represented in the text and to draw inferences exclusively from that formal but restricted interpretation.

Whether or not all meaning can be made explicit in the text is perhaps less critical than the belief that it can and that making it so is a valid scientific enterprise. This was clearly the assumption of the essayists, and it continues in our use of language for science and philosophy. Explicitness of meaning, in other words, may be better thought of as a goal rather than an achievement. But it is a goal appropriate only for the particular, specialized use of language that I have called text.

On learning a language

The contrast between language as an autonomous system for representing meaning and language as a system dependent in every case upon nonlinguistic and paralinguistic cues for the sharing of intentions—the contrast between text and utterance—applies with equal force to the problem of language acquisition. A formal theory of sentence meaning, such as Chomsky's, provides a less appropriate description of early language than would a theory of intended meanings that admitted a variety of means for realizing those intentions. Such means include a shared view of reality, a shared perceptual context, and accompanying gestures, in addition to the speech signal. At early stages of language acquisition the meaning may be specified nonlinguistically, and this meaning may then be used to break the linguistic code (Macnamara, 1972; Nelson, 1974). Language acquisition, then, is primarily a matter of learning to conventionalize more and more of the meaning in the speech signal. This is not a sudden achievement. If an utterance specifies something different from what the child is entertaining, the sentence will often be misinterpreted (Clark, 1973; Donaldson & Lloyd, 1974). But language development is not simply a matter of progressively elaborating the oral mother tongue as a means of sharing intentions. The developmental hypothesis offered here is that the ability to assign a meaning to the sentence per se, independent of its nonlinguistic interpretive context, is achieved only well into the school years. It is a complex achievement to differentiate and operate upon either what is actually said, the sentence meaning, or what is meant, the speaker's meaning. Children are relatively quick to grasp a speaker's intentions but relatively slow, I suggest, to grasp the literal meaning of what is, in fact, said.

Several studies lend plausibility to these arguments. For example, Olson and Nickerson (in press) examined the role of story or pictorial context on the detection of sentence implications. Five-year-old children were given a statement and asked if a second statement, logically related to the first, was true. For instance, they were told, "John was hit by Mary," then asked, "Did Mary hit John?" The ability of these five-year-olds to answer such a question depended on how much they knew about the characters and context mentioned in the sentences. If they did not know who John and Mary were or why the experimenter was asking the question, they could not assign a full semantic interpretation to the sentence. This and other studies suggest that children, unlike adults, assign a speaker's meaning to a simple sentence if that sentence is contextually appropriate and directly assimilable to their prior knowledge, but they have difficulty assigning a meaning to the statement alone (Carpenter & Just, 1975; Clark, 1974; Olson & Filby, 1972; Hildyard & Olson, Note 1). But by late childhood, at least among schooled children, meanings are assigned quite readily to the sentence per se. Children come to see that sentences have implications that are necessary by virtue of

sentence meaning itself. They become progressively more able to exist in a purely linguistically specified, hypothetical world for both purposes of extracting logical implications of statements and of living in those worlds that, as Ricoeur (1973) notes, are opened up by texts. This, however, is the end point of development in a literate culture and not a description of how original meanings are acquired in early language learning.

On reading

The relations between utterances and texts become acute when children are first confronted with printed books. As I have pointed out, children are familiar with using the spoken utterance as one cue among others. Children come to school with a level of oral competence in their mother-tongue only to be confronted with an exemplar of written text, the reader, which is an autonomous representation of meaning. Ideally, the printed reader depends on no cues other than linguistic cues; it represents no intentions other than those represented in the text; it is addressed to no one in particular; its author is essentially anonymous; and its meaning is precisely that represented by the sentence meaning. As a result, when children are taught to read, they are learning both to read and to treat language as text. Children familiar with the use of textlike language through hearing printed stories obviously confront less of a hurdle than those for whom both reading and that form of language are novel.

The decoding approach to reading exploits both the explicit nature of the alphabet and the explicit nature of written prose text. Ideally, since the meaning is in the text, the programmatic analysis of letters, sounds, words, and grammar would specify sentence meaning. But as I have indicated, it is precisely with sentence meaning that children have the most difficulty. Hence, the decoding of sentence meaning should be treated as the end point of development, not as the means of access to print as several writers have maintained (Reid, 1966; Richards, 1971).

On language and meaning: summary and conclusions

Clearly some aspects of meaning must be sufficiently conventionalized in the language to permit children and adults to use it as an all-purpose instrument. Thus, children must learn grammatical rules and lexical structure to use language in different contexts for different purposes. However, the degree to which this linguistic knowledge is conventionalized and formalized need not be very great in oral contexts since the listener has access to a wide range of information with which to recover the speaker's intentions. Generally, nonlinguistic cues appear to predominate in that if the speaker is elliptical or even chooses the wrong word or grammatical form, we can successfully recover the speaker's intention.

To serve the requirements of written language, however, all of the information relevant to the communication of intention must be present in the text. Further, if the text is to permit or sustain certain conclusions, as in the essayist technique, then it must become an autonomous representation of meaning. But for this purpose the meanings of the terms and the logical relations holding between them must be brought to a much higher degree of conventionalization. Words must be defined in terms of other words in the linguistic system, and rules of grammar must be specialized to make them suitable indications of the text's underlying logical structure. Once this degee of conventionalization is achieved, children or adults have sufficient basis for constructing the meaning explicitly represented by the text. Written text, I am suggesting, is largely responsible for permitting people to entertain sentence meaning per se rather than merely using the sentence as a cue to the meaning entertained by the speaker.

The differences between utterances and texts may be summarized in terms of three underlying principles: the first pertains to meaning, the second to truth, and the third to function. First, in regard to meaning, utterance and text relate in different ways to background knowledge and to the criteria for successful performance. Conventional utterances appeal for their meaning to shared experiences and interpretations, that is, to a common intuition based on shared commonsense knowledge (Lonergan, 1957; Schutz & Luckman, 1973). Utterances take for content, to use Pope's words, "What oft was tho't but ne'er so well expressed" (cited in Ong, 1971, p. 256). In most speech, as in poetry and literature, the usual reaction is assent—"How true." Statements match, in an often tantalizing way, the expectancies and experiences of the listener. Because of this appeal to expectancies, the criterion for a successful utterance is understanding on the part of the listener. The sentence is not appropriate if the listener does not comprehend. A well-formed sentence fits the requirements of the listener and, as long as this criterion is met, it does not really matter what the speaker says—"A wink is as good as a nod."

Prose text, on the other hand, appeals to premises and rules of logic for deriving implications. Whether or not the premise corresponds to common sense is irrelevant. All that is critical is that the premises are explicit and the inferences correctly drawn. The appeal is formal rather than intuitive. As a consequence, the criterion for the success of a statement in explicit prose text is its formal structure; if the text is formally adequate and the reader fails to understand, that is the reader's problem. The meaning is in the text.

Second, utterance and text appeal to different conceptions of truth. Frye (1971) has termed these underlying assumptions "truth as wisdom" and "truth as correspondence." Truth in oral utterance has to do with truth as wisdom. A statement is true if it is reasonable, plausible, and, as we have seen, congruent with dogma or the wisdom of elders; truth is assimilability to common sense. Truth in prose text, however, has to do with the correspondence

between statements and observations. Truth drops its ties to wisdom and to values, becoming the product of the disinterested search of the scientist. True statements in text may be counter to intuition, common sense, or authority. A statement is taken to be true not because the premises from which it follows are in agreement with common sense but rather because true implications follow from it, as Russell (1940) pointed out in regard to mathematics.

Third, conversational utterance and prose text involve different alignments of the functions of language. As Austin (1962) and Halliday (1970) argue, any utterance serves at least two functions simultaneously—the rhetorical or interpersonal function and the logical or ideational function. In oral speech, the interpersonal function is primary; if a sentence is inappropriate to a particular listener, the utterance is a failure. In written text, the logical or ideational functions become primary, presumably because of the indirect relation between writer and reader. The emphasis, therefore, can shift from simple communication to truth, to "getting it right" (Olson, in press). It may be this realignment of functions in written language that brings about the greater demand for explicitness and the higher degree of conventionalization.

The bias of written language toward providing definitions, making all assumptions and premises explicit, and observing the formal rules of logic produces an instrument of considerable power for building an abstract and coherent theory of reality. The development of this explicit, formal system accounts, I have argued, for the predominant features of Western culture and for our distinctive ways of using language and our distinctive modes of thought. Yet the general theories of science and philosophy that are tied to the formal uses of text provide a poor fit to daily, ordinary, practical, and personally significant experience. Oral language with its depth of resources and its multitude of paths to the same goal, while an instrument of limited power for exploring abstract ideas, is a universal means of sharing our understanding of concrete situations and practical actions. Moreover, it is the language children bring to school. Schooling, particularly learning to read, is the critical process in the transformation of children's language from utterance to text.

Acknowledgments

An early version of this paper was presented to the Epistemics meeting at Vanderbilt University, Nashville, Tenn., in February 1974 and will be published in R. Diez-Guerrero & H. Fisher (Eds.), *Logic and Language in Personality and Society*. New York: Academic Press, in press.

I am extremely grateful to the Canada Council, the Spencer Foundation, and the Van Leer Jerusalem Foundation for their support at various stages of completing this paper. I am also indebted to the many colleagues who commented on the earlier draft, including Roy Pea, Nancy Nickerson, Angela Hildyard, Bob Bracewell, Edmund Sullivan, and Frank Smith. I would also

like to thank Mary Macri who assisted with the clerical aspects of the manuscript and Isobel Gibb, Reference Librarian at OISE, who assisted with the reference editing.

Notes

1 I am indebted to Frank Smith for pointing out that I use the phrase "the meaning is in the text" as a metaphor for describing language in which the meaning is fully conventionalized.
2 The hypothesis of autonomous meaning of sentences, that is, the assumption that the meaning is in the text, may simply reflect the presupposition that linguistics, as a discipline, is autonomous.
3 This question touches upon the important epistemological issue of the formal adequacy of the methods of science. The most common argument is that almost any important theory can be shown to be formally inadequate (see Gellner, 1975).

Reference notes

1. Hildyard, A., & Olson, D. R. *On the mental representation and matching operation of action and passive sentences by children and adults*, in preparation.
2. Bracewell, R. J. *Interpretation factors in the four-card selection task*. Paper presented to the Selection Task Conference, Trento, Italy, April 1974.
3. Hidi, S. *Effects of temporal considerations in conditional reasoning*. Paper presented at the Selection Task Conference, Trento, Italy, April 1974.

References

Anderson, R. C., & Ortony, A. On putting apples into bottles: A problem of polysemy. *Cognitive Psychology*, 1975, **7**, 167–180.

Austin, J. L. *How to do things with words*. (J. O. Urmson, Ed.). New York: Oxford University Press, 1962.

Barclay, J. R. The role of comprehension in remembering sentences. *Cognitive Psychology*, 1973, **4**, 229–254.

Bloom, L. *Language development: Form and function in emerging grammars*. Cambridge, Mass.: M.I.T. Press, 1970.

Bloomfield, L. *Linguistic aspects of science*. Chicago: University of Chicago Press, 1939.

Bluck, R. S. *Plato's life and thought*. London: Routledge & Kegan Paul, 1949.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1972, **3**, 193–209.

Bransford, J. D., & Johnson, M. K. Consideration of some problems of comprehension. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Brown, R. *A first language: The early stages*. Cambridge, Mass.: Harvard University Press, 1973.

Bruner, J. S. From communication to language: A psychological perspective. *Cognition*, 1973, **3**, 255–287.

Bruner, J. S., & Olson, D. R. Symbols and texts as the tools of intellect. In *The Psychology of the 20th Century, Vol. VII: Piaget's developmental and cognitive psychology within an extended context*. Zurich: Kindler, in press.

Buhler, K. *Sprachtheorie*. Jena, Germany: Gustav Fischer Verlag, 1934.

Carpenter, P., & Just, M. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 1975, **82**, 45–73.

Carroll, J. B., & Chall, J. S. (Eds.). *Toward a literate society*. New York: McGraw-Hill, 1975.

Chafe, W. *Meaning and the structure of language*. Chicago: University of Chicago Press, 1970.

Chall, J. S. *Learning to read: The great debate*. New York: McGraw-Hill, 1967.

Chase, S. *The power of words*. New York: Harcourt, Brace, 1954.

Chomsky, N. *Syntactic structures*. The Hague: Mouton, 1957.

Chomsky, N. *Aspects of a theory of syntax*. Cambridge, Mass.: M.I.T. Press, 1965.

Chomsky, N. *Problems of knowledge and freedom*. London: Fontana, 1972.

Clark, E. Non-linguistic strategies and the acquisition of word meanings. *Cognition*, 1973, **2**, 161–182.

Clark, H. H. Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics, Vol. 12: Linguistic and adjacent arts and sciences*. The Hague: Mouton, 1974.

Cole, M., Gay, J., Glick, J., & Sharp, D. *The cultural context of learning and thinking*. New York: Basic Books, 1971.

de Laguna, G. *Speech: Its function and development*. College Park, Md.: McGrath, 1970. (Originally published, 1927.)

Donaldson, M., & Lloyd, P. Sentences and situations: Children's judgments of match and mismatch. In F. Bresson (Ed.), *Current problems in psycholinguistics*. Paris: Editions du Centre National de la Recherche Scientifique, 1974.

Ellul, J. *The technological society*. New York: Vintage Books, 1964.

Fodor, J. A., Bever, T. G., & Garrett, M. F. *The psychology of language*. Toronto: McGraw-Hill, 1974.

Ford, W. G. *The language of disjunction*. Unpublished doctoral dissertation, University of Toronto, 1976.

Frye, N. *The critical path*. Bloomington: Indiana University Press, 1971.

Gadamer, H. G. *Truth and method*. New York: Seabury Press, 1975.

Gelb, I. J. *A study of writing*. Toronto: University of Toronto Press, 1952.

Gellner, E. Book review of *Against Method* by P. Feyerabend. *British Journal for the Philosophy of Science*, 1975, **26**, 331–342.

Gibson, E. J., & Levin, H. *The psychology of reading*. Cambridge, Mass.: M.I.T. Press, 1975.

Gombrich, E. The visual image. In D. R. Olson (Ed.), *Media and symbols: The forms of expression, communication and education*. (The 73rd Yearbook of the National Society for the Study of Education). Chicago: University of Chicago Press, 1974.

Goodman, K. S. Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 1967, **6**, 126–135.

Goodman, N. *Languages of art: An approach to a theory of symbols*. Indianapolis: Bobbs-Merrill, 1968.

Goodnow, J. The nature of intelligent behavior: Questions raised by cross-cultural studies. In L. Resnick (Ed.), *New approaches to intelligence*. Potomac, Md.: Erlbaum and Associates, 1976.

Goody, J., & Watt, I. The consequences of literacy. In J. Goody (Ed.), *Literacy in traditional societies*. Cambridge, Eng.: Cambridge University Press, 1968.

Gray, M. *Song and dance man: The art of Bob Dylan*. London: Abacus, 1973.

Greenfield, P. Oral and written language: The consequences for cognitive development in Africa, the United States, and England. *Language and Speech*, 1972, **15**, 169–178.

Greenfield, P., & Bruner, J. S. Culture and cognitive growth. In D. A. Goslin (Ed.), *Handbook of socialization: Theory and research*. Chicago: Rand-McNally, 1969.

Grice, H. P. Meaning. *Philosophical Review*, 1957, **66**, 377–388.

Halliday, M. A. K. Language structure and language function. In J. Lyons (Ed.), *New horizons in linguistics*. New York: Penguin Books, 1970.

Harman, G. Deep structure as logical form. In D. Davidson & G. Harman (Eds.), *Semantics of natural language*. Dordrecht, Holland: Reidel, 1972.

Havelock, E. *Preface to Plato*. Cambridge, Mass.: Harvard University Press, 1963.

Havelock, E. Prologue to Greek literacy. *Lectures in memory of Louise Tatt Semple, second series, 1966–1971*. Cincinnati: University of Oklahoma Press for the University of Cincinnati Press, 1973.

Havelock, E. *Origins of western literacy*. Toronto: Ontario Institute for Studies in Education, 1976.

Hayakawa, S. I. *Language in thought and action*. London: Allen and Unwin, 1952.

Henle, M. On the relation between logic and thinking. *Psychological Review*, 1962, **63**, 366–378.

Inhelder, B., & Piaget, J. *The growth of logical thinking*. New York: Basic Books, 1958.

Innis, H. *The bias of communication*. Toronto: University of Toronto Press, 1951.

Korzybski, A. *Science and sanity: An introduction to non-Aristotelian systems and general semantics*. Lancaster, Pa.: Science Press, 1933.

Kneale, W., & Kneale, M. *The development of logic*. Oxford: Clarendon Press, 1962.

Lakoff, G. Linguistics and natural logic. In D. Davidson & G. Harman (Eds.), *Semantics of natural language*. Dordrecht, Holland: Reidel, 1972.

Locke, J. *An essay concerning human understanding*. (J. W. Yolton, Ed.). London: Dent, 1961. (Originally published, 1690.)

Locke, J. *Some thoughts concerning education*. (Introduction and Notes by R. H. Quick). Cambridge, Eng.: Cambridge University Press, 1880.

Lonergan, B. J. F. *Insight: A study of human understanding*. New York: Philosophical Library, 1957.

Lord, A. B. *The singer of tales* (Harvard Studies in Comparative Literature, 24). Cambridge, Mass.: Harvard University Press, 1960.

Lyons, J. *Introduction to theoretical linguistics*. Cambridge, Eng.: Cambridge University Press, 1969.

Macnamara, J. The cognitive basis of language learning in infants. *Psychological Review*, 1972, **79**, 1–13.

Maimonides, M. [*Guide of the perplexed*] (S. Pines, trans.). Chicago: University of Chicago Press, 1963.

McLuhan, M. *The Gutenberg galaxy*. Toronto: University of Toronto Press, 1962.

McLuhan, M. *Understanding media: The extensions of man*. Toronto: McGraw-Hill, 1964.

McNeill, D. *The acquisition of language*. New York: Harper & Row, 1970.

Neimark, E. D., & Slotnick, N. S. Development of the understanding of logical connectives. *Journal of Educational Psychology*, 1970, **61**, 451–460.

Nelson, K. Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, 1974, **81**, 267–285.

Olson, D. R. The languages of instruction. In R. Spiro (Ed.), *Schooling and the acquisition of knowledge*. Potomac, Md.: Erlbaum and Associates, in press.

Olson, D. R., & Filby, N. On the comprehension of active and passive sentences. *Cognitive Psychology*, 1972, **3**, 361–381.

Olson, D. R., & Nickerson, N. The contexts of comprehension: Children's inability to draw implications from active and passive sentences. *Journal of Experimental Child Psychology*, in press.

Ong, W. J. *Ramus, method and the decay of dialogue*. Cambridge, Mass.: Harvard University Press, 1958. (Reprinted by Octagon Books, 1974.)

Ong, W. J. *Rhetoric, romance and technology: Studies in the interaction of expression and culture*. Ithaca: Cornell University Press, 1971.

Paris, S. G., & Carter, A. Y. Semantic and constructive aspects of sentence memory in children. *Developmental Psychology*, 1973, **9**, 109–113.

Parry, M. The making of Homeric verse. In A. Parry (Ed.), *The collected papers of Milman Parry*. Oxford: Clarendon Press, 1971.

Piaget, J. Intellectual evolution from adolescence to adulthood. *Human Development*, 1972, **15**, 1–12.

Pike, R., & Olson, D. R. A question of more or less. *Child Development*, in press.

Popper, K. *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press, 1972.

Reid, J. F. Learning to think about reading. *Educational Research*, 1966, **9**, 56–62.

Richards, I. A. Instructional engineering. In S. Baker, J. Barzun, & I. A. Richards (Eds.), *The written word*. Rowley, Mass.: Newbury House, 1971.

Ricoeur, P. Creativity in language: Word, polysemy and metaphor. *Philosophy Today*, 1973, **17**, 97–111.

Russell, B. *An inquiry into meaning and truth*. London: Allen and Unwin, 1940.

Scribner, S., & Cole, M. Cognitive consequences of formal and informal education. *Science*, 1973, **182**, 553–559.

Schutz, A., & Luckmann, T. [*The structures of the life world*] (R. Zaner, & H. Engelhardt, trans.) Evanston, Ill.: Northwestern University Press, 1973.

Smith, F. *Comprehension and learning*. Toronto: Holt, Rinehart & Winston, 1975.

Sprat, T. *History of the Royal Society of London for the improving of natural knowledge*. (J. I. Cope and H. W. Jones, Eds.). St. Louis: Washington University Press, 1966. (Originally published, London, 1667.)

Staudenmayer, H. Understanding conditional reasoning with meaningful propositions. In R. J. Falmagne (Ed.), *Reasoning, representation and process*. Hillsdale, N.J.: Erlbaum and Associates, 1975.

Strawson, P. F. *Meaning and truth: An inaugural lecture delivered before the University of Oxford*. Oxford: Clarendon Press, 1970.

Suppes, P., & Feldman, S. Young children's comprehension of logical connectives. *Journal of Experimental Child Psychology*, 1971, **12**, 304–317.

Taplin, J. E., & Staudenmayer, H. Interpretation of abstract conditional sentences in deductive reasoning. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 530–542.

Wason, P. C., & Johnson-Laird, P. N. A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, 1970, **61**, 509–515.

Wason, P. C., & Johnson-Laird, P. N. *The psychology of reasoning*. London: B. T. Batsford, 1972.

62

WHAT NO BEDTIME STORY MEANS

Narrative skills at home and school*

S. B. Heath

"Ways of taking" from books are a part of culture and as such are more varied than current dichotomies between oral and literate traditions and relational and analytic cognitive styles would suggest. Patterns of language use related to books are studied in three literate communities in the Southeastern United States, focusing on such "literacy events" as bedtime story reading. One community, Maintown, represents mainstream, middle-class school-oriented culture; Roadville is a white mill community of Appalachian origin; the third, Trackton, is a black mill community of recent rural origin. The three communities differ strikingly in their patterns of language use and in the paths of language socialization of their children. Trackton and Roadville are as different from each other as either is from Maintown, and the differences in preschoolers' language use are reflected in three different patterns of adjustment to school. This comparative study shows the inadequacy of the prevalent dichotomy between oral and literate traditions, and points also to the inadequacy of unilinear models of child language development and dichotomies between types of cognitive styles. Study of the development of language use in relation to written materials in home and community requires a broad framework of sociocultural analysis. (Cross-cultural analysis, ethnography of communication, language development, literacy, narratives.)

In the preface to *S/Z*, Roland Barthes' work on ways in which readers read, Richard Howard writes: "We require an education in literature . . . in order to discover that *what we have assumed* – with the complicity of our teachers – *was nature is in fact culture, that what was given is no more than a way of taking*" (emphasis not in the original; Howard 1974: ix).[1] This statement reminds us that the *culture* children learn as they grow up is, in fact, "ways

Source: *Language in Society*, 1982, 11, 49–76.

of taking" meaning from the environment around them. The means of making sense from books and relating their contents to knowledge about the real world is but one "way of taking" that is often interpreted as "natural" rather than learned. The quote also reminds us that teachers (and researchers alike) have not recognized that ways of taking from books are as much a part of learned behavior as are ways of eating, sitting, playing games, and building houses.

As school-oriented parents and their children interact in the pre-school years, adults give their children, through modeling and specific instruction, ways of taking from books which seem natural in school and in numerous institutional settings such as banks, post offices, businesses, or government offices. These *mainstream* ways exist in societies around the world that rely on formal educational systems to prepare children for participation in settings involving literacy. In some communities these ways of schools and institutions are very similar to the ways learned at home; in other communities the ways of school are merely an overlay on the home-taught ways and may be in conflict with them.[2]

Yet little is actually known about what goes on in story-reading and other literacy-related interactions between adults and preschoolers in communities around the world. Specifically, though there are numerous diary accounts and experimental studies of the preschool reading experiences of mainstream middle-class children, we know little about the specific literacy features of the environment upon which the school expects to draw. Just how does what is frequently termed "the literate tradition" envelope the child in knowledge about interrelationships between oral and written language, between knowing something and knowing ways of labelling and displaying it? We have even less information about the variety of ways children from *non-mainstream* homes learn about reading, writing, and using oral language to display knowledge in their preschool environment. The general view has been that whatever it is that mainstream school-oriented homes have, these other homes do not have it; thus these children are not from the literate tradition and are not likely to succeed in school.

A key concept for the empirical study of ways of taking meaning from written sources across communities is that of *literacy events*: occasions in which written language is integral to the nature of participants' interactions and their interpretive processes and strategies. Familiar literacy events for mainstream preschoolers are bedtime stories, reading cereal boxes, stop signs, and television ads, and interpreting instructions for commercial games and toys. In such literacy events, participants follow socially established rules for verbalizing what they know from and about the written material. Each community has rules for socially interacting and sharing knowledge in literacy events.

This paper briefly summarizes the ways of taking from printed stories families teach their preschoolers in a cluster of mainstream school-oriented

neighborhoods of a city in the Southeastern region of the United States. We then describe two quite different ways of taking used in the homes of two English-speaking communities in the same region that do not follow the school-expected patterns of bookreading and reinforcement of these patterns in oral storytelling. Two assumptions underlie this paper and are treated in detail in the ethnography of these communities (Heath forthcoming b): (1) Each community's ways of taking from the printed word and using this knowledge are interdependent with the ways children learn to talk in their social interactions with caregivers. (2) There is little or no validity to the time-honored dichotomy of "the literate tradition" and "the oral tradition." This paper suggests a frame of reference for both the community patterns and the paths of development children in different communities follow in their literacy orientations.

Mainstream school-oriented bookreading

Children growing up in mainstream communities are expected to develop habits and values which attest to their membership in a "literate society." Children learn certain customs, beliefs, and skills in early enculturation experiences with written materials: the bedtime story is a major literacy event which helps set patterns of behavior that recur repeatedly through the life of mainstream children and adults.

In both popular and scholarly literature, the "bedtime story" is widely accepted as a given – a natural way for parents to interact with their child at bedtime. Commercial publishing houses, television advertising, and children's magazines make much of this familiar ritual, and many of their sales pitches are based on the assumption that in spite of the intrusion of television into many patterns of interaction between parents and children, this ritual remains. Few parents are fully conscious of what bedtime storyreading means as preparation for the kinds of learning and displays of knowledge expected in school. Ninio and Bruner (1978), in their longitudinal study of one mainstream middle-class mother–infant dyad in joint picture-book reading, strongly suggest a universal role of bookreading in the achievement of labelling by children.

In a series of "reading cycles," mother and child alternate turns in a dialogue: the mother directs the child's attention to the book and/or asks what-questions and/or labels items on the page. The items to which the what-questions are directed and labels given are two-dimensional representations of three-dimensional objects, so that the child has to resolve the conflict between perceiving these as two-dimensional objects and as representations of a three-dimensional visual setting. The child does so "by assigning a privileged, autonomous status to pictures as visual objects" (1978: 5). The arbitrariness of the picture, its decontextualization, and its existence as something which cannot be grasped and manipulated like its "real" counterparts

is learned through the routines of structured interactional dialogue in which mother and child take turns playing a labelling game. In a "scaffolding" dialogue (cf. Cazden 1979), the mother points and asks "What is x?" and the child vocalizes and/or gives a nonverbal signal of attention. The mother then provides verbal feedback and a label. Before the age of two, the child is socialized into the "initiation-reply-evaluation sequences" repeatedly described as the central structural feature of classroom lessons (e.g., Sinclair and Coulthard 1975; Griffin and Rumphry 1978; Mehan 1979). Teachers ask their students questions which have answers prespecified in the mind of the reacher. Students respond, and teacher provide feedback, usually in the form of an evaluation. Training in ways of responding to this pattern begins very early in the labelling activities of mainstream parents and children.

Maintown ways

This patterning of "incipient literacy" (Scollon and Scollon 1979) is similar in many ways to that of the families of fifteen primary-level school teachers in Maintown, a cluster of middle-class neighborhoods in a city of the Piedmont Carolinas. These families (all of whom identify themselves as "typical," "middle-class," or "mainstream,") had preschool children, and the mother in each family was either teaching in local public schools at the time of the study (early 1970s), or had taught in the academic year preceding participation in the study. Through a research dyad approach, using teacher–mothers as researchers with the ethnographer, the teacher–mothers audio-recorded their children's interactions in their primary network – mothers, fathers, grandparents, maids, siblings, and frequent visitors to the home. Children were expected to learn the following rules in literacy events in these nuclear households:

(1) As early as six months of age, children *give attention to books and information derived from books*. Their rooms contain bookcases and are decorated with murals, bedspreads, mobiles, and stuffed animals which represent characters found in books. Even when these characters have their origin in television programs, adults also provide books which either repeat or extend the characters' activities on television.
(2) Children, from the age of six months, *acknowledge questions about books*. Adults expand nonverbal responses and vocalizations from infants into fully formed grammatical sentences. When children begin to verbalize about the contents of books, adults extend their questions from simple requests for labels (What's that? Who's that?) to ask about the attributes of these items (What does the doggie say? What color is the ball?)
(3) From the time they start to talk, children *respond to conversational allusions to the content of books; they act as question-answerers who have*

a knowledge of books. For example, a fuzzy black dog on the street is likened by an adult to Blackie in a child's book: "Look, there's a Blackie. Do you think *he's* looking for a boy?" Adults strive to maintain with children a running commentary on any event or object which can be book-related, thus modelling for them the extension of familiar items and events from books to new situational contexts.

(4) Beyond two years of age, children *use their knowledge of what books do to legitimate their departures from* "*truth*." Adults encourage and reward "book talk," even when it is not directly relevant to an ongoing conversation. Children are allowed to suspend reality, to tell stories which are not true, to ascribe fiction-like features to everyday objects.

(5) Preschool children *accept book and book-related activities as entertainment*. When preschoolers are "captive audiences" (e.g., waiting in a doctor's office, putting a toy together, or preparing for bed), adults reach for books. If there are no books present, they talk about other objects as though they were pictures in books. For example, adults point to items, and ask children to name, describe, and compare them to familiar objects in their environment. Adults often ask children to state their likes or dislikes, their view of events, and so forth, at the end of the captive audience period. These affective questions often take place while the next activity is already underway (e.g., moving toward the doctor's office, putting the new toy away, or being tucked into bed), and adults do not insist on answers.

(6) Preschoolers *announce their own factual and fictive narratives* unless they are given in response to direct adult elicitation. Adults judge as most acceptable those narratives which open by orienting the listener to setting and main character. Narratives which are fictional are usually marked by formulaic openings, a particular prosody, or the borrowing of episodes in story books.

(7) When children are about three years old, adults discourage the highly interactive participative role in bookreading children have hitherto played and children *listen and wait as an audience*. No longer does either adult or child repeatedly break into the story with questions and comments. Instead, children must listen, store what they hear, and on cue from the adult, answer a question. Thus, children begin to formulate "practice" questions as they wait for the break and the expected formulaic-type questions from the adult. It is at this stage that children often choose to "read" to adults rather than to be read to.

A pervasive pattern of all these features is the authority which books and book-related activities have in the lives of both the preschoolers and members of their primary network. Any initiation of a literacy event by a preschooler makes an interruption, an untruth, a diverting of attention from the matter at hand (whether it be an uneaten plate of food, a messy room,

or an avoidance of going to bed) acceptable. Adults jump at openings their children give them for pursuing talk about books and reading.

In this study, writing was found to be somewhat less acceptable as an "anytime activity," since adults have rigid rules about times, places, and materials for writing. The only restrictions on bookreading concern taking good care of books: they should not be wet, torn, drawn on, or lost. In their talk to children about books, and in their explanations of why they buy children's books, adults link school success to "learning to love books," "learning what books can do for you," and "learning to entertain yourself and to work independently." Many of the adults also openly expressed a fascination with children's books "nowadays." They generally judged them as more diverse, wide-ranging, challenging, and exciting than books they had as children.

The mainstream pattern

A close look at the way bedtime story routines in Maintown taught children how to take meaning from books raises a heavy sense of the familiar in all of us who have acquired mainstream habits and values. Throughout a lifetime, any school-successful individual moves through the same processes described above thousands of times. Reading for comprehension involves an internal replaying of the same types of questions adults ask children of bedtime stories. We seek *what-explanations*, asking what the topic is, establishing it as predictable and recognizing it in new situational contexts by classifying and categorizing it in our mind with other phenomena. The what-explanation is replayed in learning to pick out topic sentences, write outlines, and answer standardized tests which ask for the correct titles to stories, and so on. In learning to read in school, children move through a sequence of skills designed to teach what-explanations. There is a tight linear order of instruction which recapitulates the bedtime story pattern of breaking down the story into small bits of information and teaching children to handle sets of related skills in isolated sequential hierarchies.

In each individual reading episode in the primary years of schooling, children must move through what-explanations before they can provide *reason-explanations* or *affective commentaries*. Questions about why a particular event occurred or why a specific action was right or wrong come at the end of primary-level reading lessons, just as they come at the end of bedtime stories. Throughout the primary grade levels, what-explanations predominate, reason-explanations come with increasing frequency in the upper grades, and affective comments most often come in the extra-credit portions of the reading workbook or at the end of the list of suggested activities in text books across grade levels. This sequence characterizes the total school career. High school freshmen who are judged poor in compositional and reading skills spend most of their time on what-explanations and practice in advanced versions of bedtime story questions and answers. They are given little or no

chance to use reason-giving explanations or assessments of the actions of stories. Reason-explanations result in configurational rather than hierarchical skills, are not predictable, and thus do not present content with a high degree of redundancy. Reason-giving explanations tend to rely on detailed knowledge of a specific domain. This detail is often unpredictable to teachers, and is not as highly valued as is knowledge which covers a particular area of knowledge with less detail but offers opportunity for extending the knowledge to larger and related concerns. For example, a primary-level student whose father owns a turkey farm may respond with reason-explanations to a story about a turkey. His knowledge is intensive and covers details perhaps not known to the teacher and not judged as relevant to the story. The knowledge is unpredictable and questions about it do not continue to repeat the common core of content knowledge of the story. Thus such configured knowledge is encouraged only for the "extras" of reading – an extra-credit oral report or a creative picture and story about turkeys. This kind of knowledge is allowed to be used once the hierarchical what-explanations have been mastered and displayed in a particular situation and, in the course of one's academic career, only when one has shown full mastery of the hierarchical skills and subsets of related skills which underlie what-explanations. Thus, reliable and successful participation in the ways of taking from books that teachers view as natural must, in the usual school way of doing things, precede other ways of taking from books.

These various ways of taking are sometimes referred to as "cognitive styles" or "learning styles." It is generally accepted in the research literature that they are influenced by early socialization experiences and correlated with such features of the society in which the child is reared as social organization, reliance on authority, male–female roles, and so on. These styles are often seen as two contrasting types, most frequently termed "field independent-field dependent" (Witkin et al. 1966) or "analytic-relational" (Kagan, Sigel, and Moss 1963; Cohen 1968, 1969, 1971). The analytic field-independent style is generally presented as that which correlates positively with high achievement and general academic and social success in school. Several studies discuss ways in which this style is played out in school – in preferred ways of responding to pictures and written text and selecting from among a choice of answers to test items.

Yet, we know little about how behaviors associated with either of the dichotomized cognitive styles (field-dependent/relational and field-independent/analytic) were learned in early patterns of socialization. To be sure, there are vast individual differences which may cause an individual to behave so as to be categorized as having one or the other of these learning styles. But much of the literature on learning styles suggests a preference for one or the other is learned in the social group in which the child is reared and in connection with other ways of behaving found in that culture. But how is a child socialized into an analytic/field-independent style? What kinds of

interactions does he enter into with his parents and the stimuli of his environment which contribute to the development of such a style of learning? How do these interactions mold selective attention practices such as "sensitivity to parts of objects," "awareness of obscure, abstract, nonobvious features," and identification of "abstractions based on the features of items" (Cohen 1969: 844–45)? Since the predominant stimuli used in school to judge the presence and extent of these selective attention practices are written materials, it is clear that the literacy orientation of preschool children is central to these questions.

The foregoing descriptions of how Maintown parents socialize their children into a literacy orientation fit closely those provided by Scollon and Scollon for their own child Rachel. Through similar practices, Rachel was "literate before she learned to read" (1979: 6). She knew, before the age of two, how to focus on a book and not on herself. Even when she told a story about herself, she moved herself out of the text and saw herself as author, as someone different from the central character of her story. She learned to pay close attention to the parts of objects, to name them, and to provide a running commentary on features of her environment. She learned to manipulate the contexts of items, her own activities, and language to achieve book-like, decontextualized, repeatable effects (such as puns). Many references in her talk were from written sources; others were modelled on stories and questions about these stories. The substance of her knowledge, as well as her ways of framing knowledge orally, derived from her familiarity with books and bookreading. No doubt, this development began by labelling in the dialogue cycles of reading (Ninio and Bruner 1978), and it will continue for Rachel in her preschool years along many of the same patterns described by Cochran-Smith (1981) for a mainstream nursery school. There teacher and students negotiated story-reading through the scaffolding of teachers' questions and running commentaries which replayed the structure and sequence of story-reading learned in their mainstream homes.

Close analyses of how mainstream school-oriented children come to learn to take from books at home suggest that such children learn not only how to take meaning from books, but also how to talk about it. In doing the latter, they repeatedly practice routines which parallel those of classroom interaction. By the time they enter school, they have had continuous experience as information-givers; they have learned how to perform in those interactions which surround literate sources throughout school. They have had years of practice in interaction situations that are the heart of reading – both learning to read and reading to learn in school. They have developed habits of performing which enable them to run through the hierarchy of preferred knowledge about a literate source and the appropriate sequence of skills to be displayed in showing knowledge of a subject. They have developed ways of decontextualizing and surrounding with explanatory prose the knowledge gained from selective attention to objects.

They have learned to listen, waiting for the appropriate cue which signals it is their turn to show off this knowledge. They have learned the rules for getting certain services from parents (or teachers) in the reading interaction (Merritt 1979). In nursery school, they continue to practice these interaction patterns in a group rather than in a dyadic situation. There they learn additional signals and behaviors necessary for getting a turn in a group, and responding to a central reader and to a set of centrally defined reading tasks. In short, most of their waking hours during the preschool years have enculturated them into: (1) all those habits associated with what-explanations, (2) selective attention to items of the written text, *and* (3) appropriate interactional styles for orally displaying all the know-how of their literate orientation to the environment. This learning has been finely tuned and its habits are highly interdependent. Patterns of behaviors learned in one setting or at one stage reappear again and again as these children learn to use oral and written language in literacy events and to bring their knowledge to bear in school-acceptable ways.

Alternative patterns of literacy events

But what corresponds to the mainstream pattern of learning in communities that do not have this finely tuned, consistent, repetitive, and continuous pattern of training? Are there ways of behaving which achieve other social and cognitive aims in other sociocultural groups?

The data below are summarized from an ethnography of two communities – Roadville and Trackton – located only a few miles from Maintown's neighborhoods in the Piedmont Carolinas. Roadville is a white working-class community of families steeped for four generations in the life of the textile mill. Trackton is a working-class black community whose older generations have been brought up on the land, either farming their own land or working for other landowners. However, in the past decade, they have found work in the textile mills. Children of both communities are unsuccessful in school; yet both communities place a high value on success in school, believing earnestly in the personal and vocational rewards school can bring and urging their children "to get ahead" by doing well in school. Both Roadville and Trackton are literate communities in the sense that the residents of each are able to read printed and written materials in their daily lives, and on occasion they produce written messages as part of the total pattern of communication in the community. In both communities, children go to school with certain expectancies of print and, in Trackton especially, children have a keen sense that reading is something one does to learn something one needs to know (Heath 1980). In both groups, residents turn from spoken to written uses of language and vice versa as the occasion demands, and the two modes of expression seem to supplement and reinforce each other. Nonetheless there are radical differences between the two communities in

the ways in which children and adults interact in the preschool years; each of the two communities also differs from Maintown. Roadville and Trackton view children's learning of language from two radically different perspectives: in Trackton, children "learn to talk," in Roadville, adults "teach them how to talk."

Roadville

In Roadville, babies are brought home from the hospital to rooms decorated with colorful, mechanical, musical, and literacy-based stimuli. The walls are decorated with pictures based on nursery rhymes, and from an early age, children are held and prompted to "see" the wall decorations. Adults recite nursery rhymes as they twirl the mobile made of nursery-rhyme characters. The items of the child's environment promote exploration of colors, shapes, and textures: a stuffed ball with sections of fabrics of different colors and textures is in the crib; stuffed animals vary in texture, size, and shape. Neighbors, friends from church, and relatives come to visit and talk to the baby, and about him to those who will listen. The baby is fictionalized in the talk to him: "But this baby wants to go to sleep, doesn't he? Yes, see those little eyes gettin' heavy." As the child grows older, adults pounce on word-like sounds and turn them into "words," repeating the "words," and expanding them into well-formed sentences. Before they can talk, children are introduced to visitors and prompted to provide all the expected politeness formulas, such as "Bye-bye," "Thank you," and so forth. As soon as they can talk, children are reminded about these formulas, and book or television characters known to be "polite" are involved as reinforcement.

In each Roadville home, preschoolers first have cloth books, featuring a single object on each page. They later acquire books which provide sounds, smells, and different textures or opportunities for practicing small motor skills (closing zippers, buttoning buttons, etc.). A typical collection for a two-year-old consisted of a dozen or so books – eight featured either the alphabet or numbers, others were books of nursery rhymes, simplified Bible stories, or "real-life" stories about boys and girls (usually taking care of their pets or exploring a particular feature of their environment). Books based on Sesame Street characters were favorite gifts for three- and four-year-olds.

Reading and reading-related activities occur most frequently before naps or at bedtime in the evening. Occasionally an adult or older child will read to a fussy child while the mother prepares dinner or changes a bed. On weekends, fathers sometimes read with their children for brief periods of time, but they generally prefer to play games or play with the children's toys in their interactions. The following episode illustrates the language and social interactional aspects of these bedtime events; the episode takes place between Wendy (2;3 at the time of this episode) and Aunt Sue who is putting her to bed.

[Aunt Sue (AS) picks up book, while Wendy (W) crawls about the floor, ostensibly looking for something]

W: uh uh

AS: Wendy, we're gonna read, uh, read this story, come on, hop up here on this bed.

[Wendy climbs up on the bed, sits on top of the pillow, and picks up her teddy bear]

[Aunt Sue opens book, points to puppy]

AS: Do you remember what this book is about? See the puppy? What does the puppy do?

[Wendy plays with the bear, glancing occasionally at pages of the book, as Aunt Sue turns. Wendy seems to be waiting for something in the book]

AS: See the puppy?

[Aunt Sue points to the puppy in the book and looks at Wendy to see if she is watching]

W: uh huh, yea, yes ma'am

AS: Puppy sees the ant, he's a li'l

[Wendy drops the bear and turns to book.]

fellow. Can you see that ant? Puppy has a little ball.

W: ant bite puppy

[Wendy points to ant, pushing hard on the book]

AS: No, the ant won't bite the puppy, the [turns page] puppy wants to play with the ant, see?

[Wendy tries to turn the page back; AS won't let her, and Wendy starts to squirm and russ]

AS: Look here, here's someone else, the puppy

[Wendy climbs down off the bed and gets another book]

W: read this one

AS: Okay, you get back up here now. [Wendy gets back on bed]

AS: This book is your ABC book. See the A, look, here, on your spread, there's an A. You find the A. [The second book is a cloth book, old and tattered, and long a favorite of Wendy's. It features an apple on the cover, and its front page has an ABC block and ball. Through the book, there is a single item on each page, with a large representation of the first letter of the word commonly used to name the item. As AS turns the page, Wendy begins to crawl about on her quilt, which shows ABC blocks interspersed with balls and apples. Wendy points to each of the A's on the blanket and begins talking to herself. AS reads the book, looks up, and sees Wendy pointing to the A's in her quilt.]

AS: That's an A, can you find the A on your blanket?

W: there it is, this one, there's the hole too. [pokes her finger through a place where the threads have broken in the quilting]
AS: [AS points to ball in book] Stop that, find the ball, see, here's another ball.

This episode characterizes the early orientation of Roadville children to the written word. Bookreading time focuses on letters of the alphabet, numbers, names of basic items pictured in books, and simplified retellings of stories in the words of the adult. If the content or story plot seems too complicated for the child, the adult tells the story in short, simple sentences, frequently laced with requests that the child give what-explanations.

Wendy's favorite books are those with which she can participate: that is, those to which she can answer, provide labels, point to items, give animal sounds, and "read" the material back to anyone who will listen to her. She memorizes the passages and often knows when to turn the pages to show that she is "reading." She holds the book in her lap, starts at the beginning, and often reads the title. "Puppy."

Adults and children use either the title of the book or phrases such as "the book about a puppy" to refer to reading material. When Wendy acquires a new book, adults introduce the book with phrases such as "This is a book about a duck, a little yellow duck. See the duck. Duck goes quack quack." On introducing a book, adults sometimes ask the child to recall when they have seen a "real" specimen such as that one treated in the book: "Remember the duck on the College lake?" The child often shows no sign of linking the yellow fluffy duck in the book with the large brown and grey mallards on the lake, and the adult makes no efforts to explain that two such disparate looking objects go by the same name.

As Wendy grows older, she wants to "talk" during the long stories, Bible stories, and carry out the participation she so enjoyed with the alphabet books. However, by the time she reaches three and a half, Wendy is restrained from such wide-ranging participation. When she interrupts, she is told:

> Wendy, stop that, you be quiet when someone is reading to you. You listen; now sit still and be quiet.

Often Wendy immediately gets down and runs away into the next room saying "no, no." When this happens, her father goes to get her, pats her bottom, and puts her down hard on the sofa beside him. "Now you're gonna learn to listen." During the third and fourth years, this pattern occurs more and more frequently; only when Wendy can capture an aunt who does not visit often does she bring out the old books and participate with them. Otherwise, parents, Aunt Sue, and other adults insist that she be read a story and that she "listen" quietly.

When Wendy and her parents watch television, eat cereal, visit the grocery store, or go to church, adults point out and talk about many types of written material. On the way to the grocery, Wendy (3;8) sits in the backseat, and when her mother stops at a corner, Wendy says "Stop." Her mother says "Yes, that's a stop sign." Wendy has, however, misread a yield sign as *stop*. Her mother offers no explanation of what the actual message on the sign is, yet when she comes to the sign, she stops to yield to an oncoming car. Her mother, when asked why she had not given Wendy the word "yield," said it was too hard, Wendy would not understand, and "it's not a word we use like *stop*."

Wendy recognized animal cracker boxes as early as 10 months, and later, as her mother began buying other varieties, Wendy would see the box in the grocery store and yell "Cook cook." Her mother would say, "Yes, those are cookies. Does Wendy want a cookie?" One day Wendy saw a new type of cracker box, and screeched "Cook cook." Her father opened the box and gave Wendy a cracker and waited for her reaction. She started the "cookie," then took it to her mother, saying "You eat." The mother joined in the game and said "Don't you want your *cookie*?" Wendy said "No cookie. You eat." "But Wendy, it's a cookie box, see?", and her mother pointed to the C of *crackers* on the box. Wendy paid no attention and ran off into another room.

In Roadville's literacy events, the rules for cooperative discourse around print are repeatedly practiced, coached, and rewarded in the preschool years. Adults in Roadville believe that instilling in children the proper use of words and understanding of the meaning of the written word are important for both their educational and religious success. Adults repeat aspects of the learning of literacy events they have known as children. In the words of one Roadville parent: "It was then that I began to learn . . . when my daddy kept insisting I *read* it, *say* it right. It was then that I *did* right, in his view."

The path of development for such performance can be described in three overlapping stages. In the first, children are introduced to discrete bits and pieces of books – separate items, letters of the alphabet, shapes, colors, and commonly represented items in books for children (apple, baby, ball, etc.). The latter are usually decontextualized, not pictured in their ordinary contexts, and they are represented in two-dimensional flat line drawings. During this stage, children must participate as predictable information-givers and respond to questions that ask for specific and discrete bits of information about the written matter. In these literacy events, specific features of the two-dimensional items in books which are different from their "real" counterparts are not pointed out. A ball in a book is flat; a duck in a book is yellow and fluffy; trucks, cars, dogs, and trees talk in books. No mention is made of the fact that such features do not fit these objects in reality. Children are not encouraged to move their understanding of books into other situational contexts or to apply it in their general knowledge of the world about them.

In the second stage, adults demand an acceptance of the power of print to entertain, inform, and instruct. When Wendy could no longer participate by contributing her knowledge at any point in the literacy event, she learned to recognize bookreading as a performance. The adult exhibited the book to Wendy: she was to be entertained, to learn from the information conveyed in the material, and to remember the book's content for the sequential followup questioning, as opposed to ongoing cooperative participatory questions.

In the third stage, Wendy was introduced to preschool workbooks which provided story information and was asked questions or provided exercises and games based on the content of the stories or pictures. Follow-the-number coloring books and preschool "push-out and paste" workbooks on shapes, colors, and letters of the alphabet reinforced repeatedly that the written word could be taken apart into small pieces and one item linked to another by following rules. She had practice in the linear, sequential nature of books: begin at the beginning, stay in the lines for coloring, draw straight lines to link one item to another, write your answers on lines, keep your letters straight, match the cutout letter to diagrams of letter shapes.

The differences between Roadville and Maintown are substantial. Roadville adults do not extend either the content or the habits of literacy events beyond bookreading. They do not, upon seeing an item or event in the real world, remind children of a similar event in a book and launch a running commentary on similarities and differences. When a game is played or a chore done, adults do not use literate sources. Mothers cook without written recipes most of the time; if they use a recipe from a written source, they do so usually only after confirmation and alteration by friends who have tried the recipe. Directions to games are read, but not carefully followed, and they are not talked about in a series of questions and answers which try to establish their meaning. Instead, in the putting together of toys or the playing of games, the abilities or preferences of one party prevail. For example, if an adult knows how to put a toy together, he does so; he does not talk about the process, refer to the written material and "translate" for the child, or try to sequence steps so the child can do it.[3] Adults do not talk about the steps and procedures of *how* to do things; if a father wants his preschooler to learn to hold a miniature bat or throw a ball, he says "Do it this way." He does not break up "this way" into such steps as "Put your fingers around here," "Keep your thumb in this position," "Never hold it above this line." Over and over again, adults do a task and children observe and try it, being reinforced only by commands such as "Do it like this," "Watch that thumb."

Adults at tasks do not provide a running verbal commentary on what they are doing. They do not draw the attention of the child to specific features of the sequences of skills or the attributes of items. They do not ask questions of the child, except questions which are directive or scolding in nature, ("Did you bring the ball?" "Didn't you hear what I said?"). Many of their

commands contain idioms which are not explained: "Put it up," or "Put that away now" (meaning to put it in the place where it usually belongs), or "Loosen up," said to a four-year-old boy trying to learn to bat a ball. Explanations which move beyond the listing of names of items and their features are rarely offered by adults. Children do not ask questions of the type "But I don't understand. What is that?" They appear willing to keep trying, and if there is ambiguity in a set of commands, they ask a question such as "You want me to do this?" (demonstrating their current efforts), or they try to find a way of diverting attention from the task at hand.

Both boys and girls during their preschool years are included in many adult activities, ranging from going to church to fishing and camping. They spend a lot of time observing and asking for turns to try specific tasks, such as putting a worm on the hook or cutting cookies. Sometimes adults say "No, you're not old enough." But if they agree to the child's attempt at the task, they watch and give directives and evaluations: "That's right, don't twist the cutter." "Turn like this." "Don't try to scrape it up now, let me do that." Talk about the task does not segment its skills and identify them, nor does it link the particular task or item at hand to other tasks. Reason-explanations such as "If you twist the cutter, the cookies will be rough on the edge," are rarely given, or asked for.

Neither Roadville adults nor children shift the context of items in their talk. They do not tell stories which fictionalize themselves or familiar events. They reject Sunday School materials which attempt to translate Biblical events into a modern-day setting. In Roadville, a story must be invited or announced by someone other than the storyteller, and only certain community members are designated good storytellers. A story is recognized by the group as a story about one and all. It is a true story, an actual event which occurred to either the storyteller or to someone else present. The marked behavior of the storyteller and audience alike is seen as exemplifying the weaknesses of all and the need for persistence in overcoming such weaknesses. The sources of stories are personal experience. They are tales of transgressions which make the point of reiterating the expected norms of behavior of man, woman, fisherman, worker, and Christian. They are true to the facts of the event.

Roadville parents provide their children with books; they read to them and ask questions about the books' contents. They choose books which emphasize nursery rhymes, alphabet learning, animals, and simplified Bible stories, and they require their children to repeat from these books and to answer formulaic questions about their contents. Roadville adults also ask questions about oral stories which have a point relevant to some marked behavior of a child. They use proverbs and summary statements to remind their children of stories and to call on them for simple comparisons of the stories' contents to their own situations. Roadville parents coach children in their telling of a story, forcing them to tell about an incident as it has been

pre-composed or pre-scripted in the head of the adult. Thus, in Roadville, children come to know a story as either an accounting from a book, or a factual account of a real event in which some type of marked behavior occurred and there is a lesson to be learned. Any fictionalized account of a real event is viewed as a *lie*; reality is better than fiction. Roadville's church and community life admit no story other than that which meets the definition internal to the group. Thus children cannot decontextualize their knowledge or fictionalize events known to them and shift them about into other frames.

When these children go to school they perform well in the initial stages of each of the three early grades. They often know portions of the alphabet, some colors and numbers, can recognize their names, and tell someone their address and their parents' names. They will sit still and listen to a story, and they know how to answer questions asking for what-explanations. They do well in reading workbook exercises which ask for identification of specific portions of words, items from the story, or the linking of two items, letters, or parts of words on the same page. When the teacher reaches the end of story-reading or the reading circle and asks questions such as "What did you like about the story?", relatively few Roadville children answer. If asked questions such as "What would you have done if you had been Billy [a story's main character]?", Roadville children most frequently say "I don't know" or shrug their shoulders.

Near the end of each year, and increasingly as they move through the early primary grades, Roadville children can handle successfully the initial stages of lessons. But when they move ahead to extra-credit items or to activities considered more advanced and requiring more independence, they are stumped. They turn frequently to teachers asking "Do you want me to do this? What do I do here?" If asked to write a creative story or tell it into a tape recorder, they retell stories from books; they do not create their own. They rarely provide emotional or personal commentary on their accounting of real events or book stories. They are rarely able to take knowledge learned in one context and shift it to another; they do not compare two items or events and point out similarities and differences. They find it difficult either to hold one feature of an event constant and shift all others or to hold all features constant but one. For example, they are puzzled by questions such as "What would have happened if Billy had not told the policemen what happened?" They do not know how to move events or items out of a given frame. To a question such as "What habits of the Hopi Indians might they be able to take with them when they move to a city?", they provide lists of features of life of the Hopi on the reservation. They do not take these items, consider their appropriateness in an urban setting, and evaluate the hypothetical outcome. In general, they find this type of question impossible to answer, and they do not know how to ask teachers to help them take apart the questions to figure out the answers. Thus their initial successes in reading,

being good students, following orders, and adhering to school norms of participating in lessons begin to fall away rapidly about the time they enter the fourth grade. As the importance and frequency of questions and reading habits with which they are familiar decline in the higher grades, they have no way of keeping up or of seeking help in learning what it is they do not even know they don't know.

Trackton

Babies in Trackton come home from the hospital to an environment which is almost entirely human. There are no cribs, car beds, or car seats, and only an occasional high chair or infant seat. Infants are held during their waking hours, occasionally while they sleep, and they usually sleep in the bed with parents until they are about two years of age. They are held, their faces fondled, their cheeks pinched, and they eat and sleep in the midst of human talk and noise from the television, stereo, and radio. Encapsuled in an almost totally human world, they are in the midst of constant human communication, verbal and nonverbal. They literally feel the body signals of shifts in emotion of those who hold them almost continuously; they are talked about and kept in the midst of talk about topics that range over any subject. As children make cooing or babbling sounds, adults refer to this as "noise," and no attempt is made to interpret these sounds as words or communicative attempts on the part of the baby. Adults believe they should not have to depend on their babies to tell them what they need or when they are uncomfortable; adults know, children only "come to know."

When a child can crawl and move about on his own, he plays with the household objects deemed safe for him – pot lids, spoons, plastic food containers. Only at Christmastime are there special toys for very young children; these are usually trucks, balls, doll babies, or plastic cars, but rarely blocks, puzzles, or books. As children become completely mobile, they demand ride toys or electronic and mechanical toys they see on television. They never request nor do they receive manipulative toys, such as puzzles, blocks, take-apart toys or literacy-based items, such as books or letter games.

Adults read newspapers, mail, calendars, circulars (political and civic-events related), school materials sent home to parents, brochures advertising new cars, television sets, or other products, and the Bible and other church-related materials. There are no reading materials especially for children (with the exception of children's Sunday School materials), and adults do not sit and read to children. Since children are usually left to sleep whenever and wherever they fall asleep, there is no bedtime or naptime as such. At night, they are put to bed when adults go to bed or whenever the person holding them gets tired. Thus, going to bed is not framed in any special routine. Sometimes in a play activity during the day, an older sibling will read to a younger child, but the latter soon loses interest and squirms away

to play. Older children often try to "play school" with younger children, reading to them from books and trying to ask questions about what they have read. Adults look on these efforts with amusement and do not try to convince the small child to sit still and listen.

Signs from very young children of attention to the nonverbal behaviors of others are rewarded by extra fondling, laughter, and cuddling from adults. For example, when an infant shows signs of recognizing a family member's voice on the phone by bouncing up and down in the arms of the adult who is talking on the phone, adults comment on this to others present and kiss and nudge the child. Yet when children utter sounds or combinations of sounds which could be interpreted as words, adults pay no attention. Often by the time they are twelve months old, children approximate words or phrases of adults' speech; adults respond by laughing or giving special attention to the child and crediting him with "sounding like" the person being imitated. When children learn to walk and imitate the walk of members of the community, they are rewarded by comments on their activities: "He walks just like Toby when he's tuckered out."

Children between the ages of twelve and twenty-four months often imitate the tune or "general Gestalt" (Peters 1977) of complete utterances they hear around them. They pick up and repeat chunks (usually the ends) of phrasal and clausal utterances of speakers around them. They seem to remember fragments of speech and repeat these without active production. In this first stage of language learning, the repetition stage, they imitate the intonation contours and general shaping of the utterances they repeat. Lem 1;2 in the following example illustrates this pattern.

Mother:	[talking to neighbor on porch while Lem plays with a truck on the porch nearby] But they won't call back, won't happen =
Lem:	= call back
Neighbor:	Sam's going over there Saturday, he'll pick up a form =
Lem:	= pick up on, pick up on [Lem here appears to have heard *form* as *on*]

The adults pay no attention to Lem's "talk," and their talk, in fact, often overlaps his repetitions.

In the second stage, repetition with variation, Trackton children manipulate pieces of conversation they pick up. They incorporate chunks of language from others into their own ongoing dialogue, applying productive rules, inserting new nouns and verbs for those used in the adults' chunks. They also play with rhyming patterns and varying intonation contours.

Mother:	She went to the doctor again.
Lem (2;2):	[in a sing-song fashion] went to de doctor, doctor, tractor, dis my tractor, doctor on a tractor, went to de doctor.

Lem creates a monologue, incorporating the conversation about him into his own talk as he plays. Adults pay no attention to his chatter unless it gets so noisy as to interfere with their talk.

In the third stage, participation, children begin to enter the ongoing conversations about them. They do so by attracting the adult's attention with a tug on the arm or pant leg, and they help make themselves understood by providing nonverbal reinforcements to help recreate a scene they want the listener to remember. For example, if adults are talking, and a child interrupts with seemingly unintelligible utterances, the child will make gestures, extra sounds, or act out some outstanding features of the scene he is trying to get the adult to remember. Children try to create a context, a scene, for the understanding of their utterance.

This third stage illustrates a pattern in the children's response to their environment and their ways of letting others know their knowledge of the environment. Once they are in the third stage, their communicative efforts are accepted by community members, and adults respond directly to the child, instead of talking to others about the child's activities as they have done in the past. Children continue to practice for conversational participation by playing, when alone, both parts of dialogues, imitating gestures as well as intonation patterns of adults. By 2;6 all children in the community can imitate the walk and talk of others in the community, or frequent visitors such as the man who comes around to read the gas meters. They can feign anger, sadness, fussing, remorse, silliness, or any of a wide range of expressive behaviors. They often use the same chunks of language for varying effects, depending on nonverbal support to give the language different meanings or cast it in a different key (Hymes 1974). Girls between three and four years of age take part in extraordinarily complex stepping and clapping patterns and simple repetitions of hand clap games played by older girls. From the time they are old enough to stand alone, they are encouraged in their participation by siblings and older children in the community. These games require anticipation and recognition of cues for upcoming behaviors, and the young girls learn to watch for these cues and to come in with the appropriate words and movements at the right time.

Preschool children are not asked for what-explanations of their environment. Instead, they are asked a preponderance of analogical questions which call for non-specific comparisons of one item, event, or person with another: "What's that like?" Other types of questions ask for specific information known to the child but not the adults: "Where'd you get that from?" "What do you want?" "How come you did that?" (Heath 1982). Adults explain their use of these types of questions by expressing their sense of children: they are "comers," coming into their learning by experiencing what knowing about things means. As one parent of a two-year-old boy put it: "Ain't no use me tellin' 'im: learn this, learn that, what's this, what's that? He just gotta learn, gotta know; he see one thing one place one time, he know how it go, see

sump'n like it again, maybe it be the same, maybe it won't." Children are expected to learn how to know when the form belies the meaning, and to know contexts of items and to use their understanding of these contexts to draw parallels between items and events. Parents do not believe they have a tutoring role in this learning; they provide the experiences on which the child draws and reward signs of their successfully coming to know.

Trackton children's early stories illustrate how they respond to adult views of them as "comers." The children learn to tell stories by drawing heavily on their abilities to render a context, to set a stage, and to call on the audience's power to join in the imaginative creation of story. Between the ages of two and four years, the children, in a monologue-like fashion, tell stories about things in their lives, events they see and hear, and situations in which they have been involved. They produce these spontaneously during play with other children or in the presence of adults. Sometimes they make an effort to attract the attention of listeners before they begin the story, but often they do not. Lem, playing off the edge of the porch, when he was about two and a half years of age, heard a bell in the distance. He stopped, looked at Nellie and Benjy, his older siblings, who were nearby and said:

Way
Far
Now
It a church bell
Ringin'
Dey singin'
Ringin'
You hear it?
I hear it
Far
Now.

Lem had been taken to church the previous Sunday and had been much impressed by the church bell. He had sat on his mother's lap and joined in the singing, rocking to and fro on her lap, and clapping his hands. His story, which is like a poem in its imagery and line-like prosody, is in response to the current stimulus of a distant bell. As he tells the story, he sways back and forth.

This story, somewhat longer than those usually reported from other social groups for children as young as Lem,[4] has some features which have come to characterize fully-developed narratives or stories. It recapitulates in its verbal outline the sequence of events being recalled by the storyteller. At church, the bell rang while the people sang. In the line "It a church bell," Lem provides his story's topic, and a brief summary of what is to come. This line serves a function similar to the formulae often used by older children

to open a story: "This is a story about (a church bell)." Lem gives only the slightest hint of story setting or orientation to the listener; where and when the story took place are capsuled in "Way, Far." Preschoolers in Trackton almost never hear "Once upon a time there was a ——" stories, and they rarely provide definitive orientations for their stories. They seem to assume listeners "know" the situation in which the narrative takes place. Similarly, preschoolers in Trackton do not close off their stories with formulaic endings. Lem poetically balances his opening and closing in an inclusio, beginning "Way, Far, Now." and ending "Far, Now.". The effect is one of closure, but there is no clearcut announcement of closure. Throughout the presentation of action and result of action in their stories, Trackton preschoolers invite the audience to respond or evaluate the story's actions. Lem asks "You hear it?" which may refer either to the current simulus or to yesterday's bell, since Lem does not productively use past tense endings for any verbs at this stage in his language development.

Preschool storytellers have several ways of inviting audience evaluation and interest. They may themselves express an emotional response to the story's actions; they may have another character or narrator in the story do so often using alliterative language play; or they may detail actions and results through direct discourse or sound effects and gestures. All these methods of calling attention to the story and its telling distinguish the speech event as a story, an occasion for audience and storyteller to interact pleasantly, and not simply to hear an ordinary recounting of events or actions.

Trackton children must be aggressive in inserting their stories into an ongoing stream of discourse. Storytelling is highly competitive. Everyone in a conversation may want to tell a story, so only the most aggressive wins out. The content ranges widely, and there is "truth" only in the universals of human experience. Fact is often hard to find, though it is usually the seed of the story. Trackton stories often have no point – no obvious beginning or ending; they go on as long as the audience enjoys and tolerates the storyteller's entertainment.

Trackton adults do not separate out the elements of the environment around their children to tune their attentions selectively. They do not simplify their language, focus on single-word utterances by young children, label items or features of objects in either books or the environment at large. Instead, children are continuously contextualized, presented with almost continuous communication. From this ongoing, multiple-channeled stream of stimuli, they must themselves select, practice, and determine rules of production and structuring. For language, they do so by first repeating, catching chunks of sounds, intonation contours, and practicing these without specific reinforcement or evaluation. But practice material and models are continuously available. Next the children seem to begin to sort out the productive rules for speech and practice what they hear about them with

variation. Finally, they work their way into conversations, hooking their meanings for listeners into a familiar context by recreating scenes through gestures, special sound effects, etc. These characteristics continue in their story–poems and their participation in jump-rope rhymes. Because adults do not select out, name, and describe features of the environment for the young, children must perceive situations, determine how units of the situations are related to each other, recognize these relations in other situations, and reason through what it will take to show their correlation of one situation with another. The children can answer questions such as "What's that like?" ["It's like Doug's car"] but they can rarely name the specific feature or features which make two items or events alike. For example, in the case of saying a car seen on the street is "like Doug's car," a child may be basing the analogy on the fact that this car has a flat tire and Doug's also had one last week. But the child does not name (and is not asked to name) what is alike between the two cars.

Children seem to develop connections between situations or items not by specification of labels and features in the situations, but by configuration links. Recognition of similar general shapes or patterns of links seen in one situation and connected to another, seem to be the means by which children set scenes in their nonverbal representations of individuals, and later in their verbal chunking, then segmentation and production of rules for putting together isolated units. They do not decontextualize; instead they heavily contextualize nonverbal and verbal language. They fictionalize their "true stories," but they do so by asking the audience to identify with the story through making parallels from their own experiences. When adults read, they often do so in a group. One person, reading aloud, for example, from a brochure on a new car decodes the text, displays illustrations and photographs, and listeners relate the text's meaning to their experiences asking questions and expressing opinions. Finally, the group as a whole synthesizes the written text and the negotiated oral discourse to construct a meaning for the brochure (Heath forthcoming a).

When Trackton children go to school, they face unfamiliar types of questions which ask for what-explanations. They are asked as individuals to identify items by name, and to label features such as shape, color, size, number. The stimuli to which they are to give these responses are two-dimensional flat representations which are often highly stylized and bear little resemblance to the "real" items. Trackton children generally score in the lowest percentile range on the Metropolitan Reading Readiness tests. They do not sit at their desks and complete reading workbook pages; neither do they tolerate questions about reading materials which are structured along the usual lesson format. Their contributions are in the form of "I had a duck at my house one time." "Why'd he do that?" or they imitate the sound effects teachers may produce in stories they read to the children. By the end of the first three primary grades, their general language arts scores

have been consistently low, except for those few who have begun to adapt to and adopt some of the behaviors they have had to learn in school. But the majority not only fail to learn the content of lessons, they also do not adopt the social interactional rules for school literacy events. Print in isolation bears little authority in their world. The kinds of questions asked of reading books are unfamiliar. The children's abilities to metaphorically link two events or situations and to recreate scenes are not tapped in the school; in fact, *these abilities often cause difficulties*, because they enable children to see parallels teachers did not intend, and indeed, may not recognize until the children point them out (Heath 1978).

By the end of the lessons or by the time in their total school career when reason-explanations and affective statements call for the creative comparison of two or more situations, it is too late for many Trackton children. They have not picked up along the way the composition and comprehension skills they need to translate their analogical skills into a channel teachers can accept. They seem not to know how to take meaning from reading; they do not observe the rules of linearity in writing, and their expression of themselves on paper is very limited. Orally taped stories are often much better, but these rarely count as much as written compositions. Thus, Trackton children continue to collect very low or failing grades, and many decide by the end of the sixth grade to stop trying and turn their attention to the heavy peer socialization which usually begins in these years.

From community to classroom

A recent review of trends in research on learning pointed out that "learning to read through using and learning from language has been less systematically studied than the decoding process" (Glaser 1979: 7). Put another way, how children learn to use language to read to learn has been less systematically studied than decoding skills. Learning how to take meaning from writing before one learns to read involves repeated practice in using and learning from language through appropriate participation in literacy events such as exhibitor/questioner and spectator/respondent dyads (Scollon and Scollon 1979) or group negotiation of the meaning of a written text. Children have to learn to select, hold, and retrieve content from books and other written or printed texts in accordance with their community's rules or "ways of taking," and the children's learning follows community paths of language socialization. In each society, certain kinds of childhood participation in literacy events may precede others, as the developmental sequence builds toward the whole complex of home and community behaviors characteristic of the society. The ways of taking employed in the school may in turn build directly on the preschool development, may require substantial adaptation on the part of the children, or may even run directly counter to aspects of the community's pattern.

At home

In *Maintown* homes, the construction of knowledge in the earliest preschool years depends in large part on labelling procedures and what-explanations. Maintown families, like other mainstream families, continue this kind of classification and knowledge construction throughout the child's environment and into the school years, calling it into play in response to new items in the environment and in running commentaries on old items as they compare to new ones. This pattern of linking old and new knowledge is reinforced in narrative tales which fictionalize the teller's events or recapitulate a story from a book. Thus for these children the bedtime story is simply an early link in a long chain of interrelated patterns of taking meaning from the environment. Moreover, along this chain, the focus is on the individual as respondent and cooperative negotiator of meaning from books. In particular, children learn that written language may represent not only descriptions of real events, but decontextualized logical propositions, and the occurrence of this kind of information in print or in writing legitimates a response in which one brings to the interpretation of written text selected knowledge from the real world. Moreover, readers must recognize how certain types of questions assert the priority of meanings in the written word over reality. The "real" comes into play only after prescribed decontextualized meanings; affective responses and reason-explanations follow conventional presuppositions which stand behind what-explanations.

Roadville also provides labels, features, and what-explanations, and prescribes listening and performing behaviors for preschoolers. However, Roadville adults do not carry on or sustain in continually overlapping and interdependent fashion the linking of ways of taking meaning from books to ways of relating that knowledge to other aspects of the environment. They do not encourage decontextualization; in fact, they proscribe it in their own stories about themselves and their requirements of stories from children. They do not themselves make analytic statements or assert universal truths, except those related to their religious faith. They lace their stories with synthetic (nonanalytic) statements which express, describe, and synthesize actual real-life materials. Things do not have to follow logically so long as they fit the past experience of individuals in the community. Thus children learn to look for a specific moral in stories and to expect that story to fit their facts of reality explicitly. When they themselves recount an event, they do the same, constructing the story of a real event according to coaching by adults who want to construct the story as they saw it.

Trackton is like neither Maintown nor Roadville. There are no bedtime stories; in fact, there are few occasions for reading to or with children specifically. Instead, during the time these activities would take place in mainstream and Roadville homes, Trackton children are enveloped in different kinds of social interactions. They are held, fed, talked about, and rewarded

for nonverbal, and later verbal, renderings of events they witness. Trackton adults value and respond favorably when children show they have come to know how to use language to show correspondence in function, style, configuration, and positioning between two different things or situations. Analogical questions are asked of Trackton children, although the implicit questions of structure and function these embody are never made explicit. Children do not have labels or names of attributes of items and events pointed out for them, and they are asked for reason-explanations not what-explanations. Individuals express their personal responses and recreate corresponding situations with often only a minimal adherence to the germ of truth of a story. Children come to recognize similarities of patterning, though they do not name lines, points, or items which are similar between two items or situations. They are familiar with group literacy events in which several community members orally negotiate the meaning of a written text.

At school

In the early reading stages, and in later requirements for reading to learn at more advanced stages, children from the three communities respond differently, because they have learned different methods and degrees of taking from books. In comparison to Maintown children, the habits Roadville children learned in bookreading and toy-related episodes have not continued for them through other activities and types of reinforcement in their environment. They have had less exposure to both the content of books and ways of learning from books than have mainstream children. Thus their need in schools is not necessarily for an intensification of presentation of labels, a slowing down of the sequence of introducing what-explanations in connection with bookreading. Instead they need *extension of these habits to other domains* and to opportunities for practicing habits such as producing running commentaries, creating exhibitor/questioner and spectator/respondent roles. Perhaps most important, Roadville children need to have articulated for them *distinctions in discourse strategies and structures*. Narratives of real events have certain strategies and structures; imaginary tales, flights of fantasy, and affective expressions have others. Their community's view of narrative discourse style is very narrow and demands a passive role in both creation of and response to the account of events. Moreover, these children have *to be reintroduced to a participant frame of reference to a book*. Though initially they were participants in bookreading, they have been trained into passive roles since the age of three years, and they must learn once again to be active information-givers, taking from books and linking that knowledge to other aspects of their environment.

Trackton students present an additional set of alternatives for procedures in the early primary grades. Since they usually have few of the expected "natural" skills of taking meaning from books, they must not only learn

these, but also *retain their analogical reasoning practices* for use in some of the later stages of learning to read. They must *learn to adapt the creativity in language, metaphor, fictionalization, recreation of scenes and exploration of functions and settings of items they bring to school*. These children already use narrative skills highly rewarded in the upper primary grades. They distinguish a fictionalized story from a real-life narrative. They know that telling a story can be in many ways related to play; it suspends reality, and frames an old event in a new context; it calls on audience participation to recognize the setting and participants. They must now *learn as individuals to recount factual events in a straightforward way* and *recognize appropriate occasions for reason-explanations and affective expressions*. Trackton children seem to have skipped learning to label, list features, and give what-explanations. Thus they need to *have the mainstream or school habits presented in familiar activities with explanations related to their own habits of taking meaning* from the environment. Such "simple," "natural" things as distinctions between two-dimensional and three-dimensional objects may need to be explained to help Trackton children learn the stylization and decontextualization which characterizes books.

To lay out in more specific detail how Roadville and Trackton's ways of knowing can be used along with those of mainstreamers goes beyond the scope of this paper. However, it must be admitted that a range of alternatives to ways of learning and displaying knowledge characterizes all highly school-successful adults in the advanced stages of their careers. Knowing more about how these alternatives are learned at early ages in different sociocultural conditions can help the school to provide opportunities for *all* students to avail themselves of these alternatives early in their school careers. For example, mainstream children can benefit from early exposure to Trackton's creative, highly analogical styles of telling stories and giving explanations, and they can add the Roadville true story with strict chronicity and explicit moral to their repertoire of narrative types.

In conclusion, if we want to understand the place of literacy in human societies and ways children acquire the literacy orientations of their communities, we must recognize two postulates of literacy and language development.

(1) Strict dichotomization between oral and literate traditions is a construct of researchers, not an accurate portrayal of reality across cultures.
(2) A unilinear model of development in the acquisition of language structures and uses cannot adequately account for culturally diverse ways of acquiring knowledge or developing cognitive styles.

Roadville and Trackton tell us that the mainstream type of literacy orientation is not the only type even among Western societies. They also tell us that the mainstream ways of acquiring communicative competence do not offer a universally applicable model of development. They offer proof of Hymes'

assertion a decade ago that "it is impossible to generalize validly about 'oral' vs. 'literate' cultures as uniform types" (Hymes 1973: 54).

Yet in spite of such warnings and analyses of the uses and functions of writing in the specific proposals for comparative development and organization of cultural systems (cf. Basso 1974: 432), the majority of research on literacy has focused on differences in class, amount of education, and level of civilization among groups having different literacy characteristics.

"We need, in short, a great deal of ethnography" (Hymes 1973: 57) to provide descriptions of the ways different social groups "take" knowledge from the environment. For written sources, these ways of taking may be analyzed in terms of *types of literacy events*, such as group negotiation of meaning from written texts, individual "looking things up" in reference books, writing family records in Bibles, and the dozens of other types of occasions when books or other written materials are integral to interpretation in an interaction. These must in turn be analyzed in terms of the specific *features of literacy events*, such as labelling, what-explanation, affective comments, reason-explanations, and many other possibilities. Literacy events must also be interpreted in relation to the *larger sociocultural patterns* which they may exemplify or reflect. For example, ethnography must describe literacy events in their sociocultural contexts, so we may come to understand how such patterns as time and space usage, caregiving roles, and age and sex segregation are interdependent with the types and features of literacy events a community develops. It is only on the basis of such thoroughgoing ethnography that further progress is possible toward understanding cross-cultural patterns of oral and written language uses and paths of development of communicative competence.

Notes

* One of a series of invited papers commemorating a decade of *Language in Society*.

1 First presented at the Terman Conference on Teaching at Stanford University, 1980, this paper has benefitted from cooperation with M. Cochran-Smith of the University of Pennsylvania. She shares an appreciation of the relevance of Roland Barthes' work for studies of the socialization of young children into literacy; her research (1981) on the story-reading practices of a mainstream school-oriented nursery school provides a much needed detailed account of early school orientation to literacy.

2 Terms such as *mainstream* or *middle-class* cultures or social groups are frequently used in both popular and scholarly writings without careful definition. Moreover, numerous studies of behavioral phenomena (for example, mother-child interactions in language learning) either do not specify that the subjects being described are drawn from mainstream groups or do not recognize the importance of this limitation. As a result, findings from this group are often regarded as universal. For a discussion of this problem, see Chanan and Gilchrist 1974, Payne and Bennett 1977. In general, the literature characterizes this group as school-oriented,

aspiring toward upward mobility through formal institutions, and providing enculturation which positively values routines of promptness, linearity (in habits ranging from furniture arrangement to entrance into a movie theatre), and evaluative and judgmental responses to behaviors which deviate from their norms.

In the United States, mainstream families tend to locate in neighborhoods and suburbs around cities. Their social interactions center not in their immediate neighborhoods, but around voluntary associations across the city. Thus a cluster of mainstream families (and not a community – which usually implies a specific geographic territory as the locus of a majority of social interactions) is the unit of comparison used here with the Trackton and Roadville communities.

3 Behind this discussion are findings from cross-cultural psychologists who have studied the links between verbalization of task and demonstration of skills in a hierarchical sequence, e.g., Childs and Greenfield 1980; see Goody 1979 on the use of questions in learning tasks unrelated to a familiarity with books.

4 Cf. Umiker-Sebeok's (1979) descriptions of stories of mainstream middle-class children, ages 3–5 and Sutton-Smith 1981.

References

Basso, K. (1974). The ethonography of writing. In R. Bauman & J. Sherzer (eds.), *Explorations in the ethnography of speaking*. Cambridge University Press.

Cazden, C. B. (1979). Peekaboo as an instructional model: Discourse development at home and at school. *Papers and Reports in Child Language Development* **17**: 1–29.

Chanan, G., & Gilchrist, L. (1974). *What school is for*. New York: Praeger.

Childs, C. P., & Greenfield, P. M. (1980). Informal modes of learning and teaching. In N. Warren (ed.), *Advances in cross-cultural psychology*, vol. 2 London: Academic Press.

Cochran-Smith, M. (1981). The making of a reader. Ph.D. dissertation. University of Pennsylvania.

Cohen, R. (1968). The relation between socio-conceptual styles and orientation to school requirements. *Sociology of Education* **41**: 201–20.

____. (1969). Conceptual styles, culture conflict, and nonverbal tests of intelligence. *American Anthropologist* **71** (5): 828–56.

____. (1971). The influence of conceptual rule-sets on measures of learning ability. In C. L. Brace, G. Gamble, & J. Bond (eds.), *Race and intelligence*. (Anthropological Studies, No. 8, American Anthropological Association) 41–57.

Glaser, R. (1979). Trends and research questions in psychological research on learning and schooling. *Educational Researcher* **8** (10): 6–13.

Goody, E. (1979). Towards a theory of questions. In E. N. Goody (ed.), *Questions and politeness: Strategies in social interaction*. Cambridge University Press.

Griffin, P., & Humphrey, F. (1978). Task and talk. In *The study of children's functional language and education in the early years*. Final report to the Carnegie Corporation of New York. Arlington, Va.: Center for Applied Linguistics.

Heath, S. (1978). *Teacher talk: Language in the classroom*. (Language in Education 9.) Arlington, Va.: Center for Applied Linguistics.

____. (1980). The functions and uses of literacy. *Journal of Communication* **30** (1): 123–33.

____. (1982). Questioning at home and at school: A comparative study. In G. Spindler (ed.), *Doing ethnography: Educational anthropology in action*. New York: Holt, Rinehart & Winston.

____. (forthcoming a). Protean shapes: Ever-shifting oral and literate traditions. To appear in D. Tannen (ed.), *Spoken and written language: Exploring orality and literacy*. Norwood. N.J.: Ablex.

____. (forthcoming b). *Ways with words: Ethnography of communication in communities and classrooms*.

Howard, R. (1974). A note on S/Z. In R. Barthes, *Introduction to S/Z*. Trans. Richard Miller. New York: Hill and Wang.

Hymes, D. H. (1973). On the origins and foundations of inequality among speakers. In E. Haugen & M. Bloomfield (eds.), *Language as a human problem*. New York: W. W. Norton & Co.

____. (1974). Models of the interaction of language and social life. In J. J. Gumperz & D. Hymes (eds.), *Directions in sociolinguistics*. New York: Holt, Rinehart and Winston.

Kagan, J., Sigel, I., & Moss, H. (1963). Psychological significance of styles of conceptualization. In J. Wright & J. Kagan (eds.), *Basic cognitive processes in children*. (Monographs of the society for research in child development.) **28** (2): 73–112.

Mehan, H. (1979). *Learning lessons*. Cambridge, Mass.: Harvard University Press.

Merritt, M. (1979). Service-like events during individual work time and their contribution to the nature of the rules for communication. NIE Report EP 78-0436.

Ninio, A., & Bruner, J. (1978). The achievement and antecedents of labelling. *Journal of Child Language* **5**: 1–15.

Payne, C., & Bennett, C. (1977). "Middle class aura" in public schools. *The Teacher Educator* **13** (1): 16–26.

Peters, A. (1977). Language learning strategies. *Language* **53**: 560–73.

Scollon, R., & Scollon, S. (1979). The literate two-year old: The fictionalization of self. *Working Papers in Sociolinguistics*. Austin, TX: Southwest Regional Laboratory.

Sinclair, J. M., & Coulthard, R. M. (1975). *Toward an analysis of discourse*. New York: Oxford University Press.

Sutton-Smith, B. (1981). *The folkstories of children*. Philadelphia: University of Pennsylvania Press.

Umiker-Sebeok, J. D. (1979). Preschool children's intraconversational narratives. *Journal of Child Language* **6** (1): 91–110.

Witkin, H., Faterson, F., Goodenough, R., & Birnbaum, J. (1966). Cognitive patterning in mildly retarded boys. *Child Development* **37** (2): 301–16.

63

SCHOOLING FOR LITERACY

A review of research on teacher effectiveness and school effectiveness and its implications for contemporary educational policies

D. Reynolds

An outline is given of educational policies concerned with literacy and of the teacher effectiveness and school effectiveness knowledge bases that are now central in educational discourse and which provide a specific 'technology' of practices associated with the development of literacy. It is argued that research findings concerning the 'context specificity' of effective school practices, concerning the importance of the classroom level and concerning the difficulties of getting knowledge to 'root' in schools all suggest a somewhat different orientation may be needed within the present range of educational policies concerned with literacy if they are to be effective.

Introduction

A discourse about the importance of literacy and numeracy as 'core skills' or 'basics' that all children should possess has in recent years become central to educational politics and educational policy making in the UK. Reflected in the recent suspension of requirements for primary schools to teach the full range of current national curriculum subjects at Key Stage 2 and also reflected in the current concern with so called 'basics' that has been at the heart of the incoming government's White Paper *Excellence in Schools* (Department for Education and Employment, 1997a), this discourse can be seen as reflecting a number of concerns, political, social and economic.

Firstly, it has been argued that low skill levels in the population of the UK and the associated problem of the 'trailing edge' of children leaving school with no qualifications cost British society considerably in terms of

Source: *Educational Review*, 1998, 50(2), 147–162.

lost wealth creation and cost further economically and socially in the resources required to deal with the social problems that are associated with academic failure. Although the evidence that purely educational reforms can achieve very much change in this situation is much disputed (Robinson, 1997), links between the UK's poor performance in international surveys of achievement and poor levels of economic performance have frequently been made (Reynolds & Farrell, 1996).

Secondly, it is now widely agreed that the early years of children's development, their early schooling and their levels of literacy and numeracy in Key Stages 1 and 2 are of vital importance in determining later academic and social outcomes. Primary education is seen in most research now as having larger effects than secondary education (Reynolds *et al.*, 1996a) and the 'basic' skills of literacy (reading and writing) and numeracy are now increasingly seen as determining both performance in other subject areas (like science and the humanities) at the primary stage and also academic achievement at a much later stage. Some studies indeed show correlations as high as 0.8 between children's performance on reading at age seven and their subsequent achievement scores (Sammons *et al.*, 1997).

In the specific case of literacy, in the desire to improve standards a variety of policy initiatives are being implemented. The Literacy Task Force was set up in 1996 and reported its plans in a Preliminary Report (Institute of Education, 1997) and in a Final Report entitled *The Implementation of the National Literacy Strategy* (Department for Education and Employment, 1997b), arguing for a range of initiatives to improve levels of achievement in literacy, including dedicated training days to deliver a perceived 'technology' of skills in literacy teaching to all primary teachers and additionally the 'roll out' of the existing National Literacy Project to schools where test scores on the Key Stage 2 SAT's appear to be low. Overall, the target of the present initiatives is for 80% of 11 year old pupils to achieve Level 4 or better in English by 2002.

In all this flurry of activity the international bodies of research knowledge on teacher effectiveness, school effectiveness and school improvement have increasingly come to have a central position. They feature in the various policy documents themselves: those who have contributed to these research fields have themselves increasingly become 'co-opted' as policy advisers, both formally and informally by the government. The body of knowledge is now well validated (see reviews in Reynolds & Cuttance, 1992; Reynolds *et al.*, 1996b) and is increasingly international in range (Reynolds *et al.*, 1994), generating a 'normal science' in which reviews of 'what works' are increasingly international in scope (Teddlie & Reynolds, 1998).

What we attempt to do in this paper is to survey this body of knowledge on the two key areas of teacher effectiveness and school effectiveness, assess the strength of its findings and particularly assess the extent to which knowledge of the research findings supports the range of contemporary policies

that aim at an improvement in standards of literacy. We will conclude with an assessment of the further policy developments in the general areas of school improvement and teacher development that may be necessary to deliver the goals concerning literacy standards that feature prominently in the contemporary political discourse and in contemporary educational policies.

Teacher effectiveness in literacy

There is an extensive body of knowledge concerning the behaviours of teachers who 'add value' in the area of literacy, usually measured by the use of 'pre-tests' and 'post-tests' of reading skills together with the study of the relationship between observed teacher behaviours in the classroom and the 'gain' that they produce in pupils' achievement scores over, usually, an academic year (see reviews in Scheerens, 1992; Creemers, 1994).

To take the American literature first, one of the factors to most consistently and most strongly affect reading test scores is 'opportunity-to-learn', whether it is measured as the amount of the curriculum covered or the percentage of test items taught (Brophy & Good, 1986). Opportunity-to-learn is clearly related to such factors as the length of the school day and year and to the hours of reading experience taught. It is, however, also related to the quality of teachers' classroom management and especially to what is known as 'time-on-task' (i.e. the amount of time children are actively engaged in learning activities in the classroom, as opposed to socialising, etc.). Opportunity to learn is also clearly related to the use of homework, which expands available learning time.

Another highly important factor which distinguishes effective teachers from less effective, a factor that is also connected to children's time-on-task, is the teacher's academic orientation. Effective teachers emphasise academic instruction and see learning as the main classroom goal. This means that they spend most of their time on curriculum-based learning activities and create a task-oriented, business-like, but also relaxed and supportive, environment (Brophy & Good, 1986; Griffin & Barnes, 1986; Cooney, 1994).

Obviously, the time-on-task levels of the children are strongly influenced by classroom management. Effective teachers are able to organise and manage classrooms as effective learning environments in which academic activities run smoothly, transitions (between lesson segments) are brief and little time is spent getting organised or dealing with inattention or resistance (Brophy & Good, 1986). For this to happen, good prior preparation of the classroom and the installation of clear rules and procedures (before or at the start of the school year) are essential. All in all, effective teachers manage to create a well-organised classroom with minimal disruption and misbehaviour (Evertson *et al.*, 1980; Brophy & Good, 1986; Griffin & Barnes, 1986).

Teacher expectations are also very important. Effective teachers show they believe that all children can master the curriculum (not just a percentage

of children). They emphasise the positive (e.g. if a child is not so good in one area she/he might be good in another) and these positive expectations are transmitted to children. Effective teachers emphasise the importance of effort, clarifying the relationship between effort and outcomes and helping pupils gain an internal locus of control by constantly pointing out the importance of their own work (Borich, 1996).

Research has also found that children learn more in classes where they spend time being taught or supervised by their teacher rather than working on their own. In such classes teachers spend most of their time presenting information through lecture and demonstration. Teacher-led discussion as opposed to individual work dominates. This is not to say that all individual work is negative, individual practice is even necessary and important, but many teachers have been found to rely too much on pupils working on their own, at the expense of lecture–demonstration and class discussion (Evertson *et al*., 1980).

Research has found that classrooms where more time is spent teaching the whole class, rather than on letting individual pupils work by themselves (e.g. with worksheets), see higher pupil achievement gains. This is mainly because teachers in these classrooms provide more thoughtful and thorough presentations, spend less time on classroom management, enhance time-on-task and can make more child contacts. Teachers giving whole-class instruction have also been found to spend more time monitoring children's achievement. There were also likely to be less child disruptions with this method, thus again increasing time-on-task (Evertson *et al*., 1980; Brophy, 1986; Walberg, 1986). The effective teacher carries the content personally to the student, rather than relying on curriculum material or textbooks to do so.

This focus on the teacher presenting material in an active way to students should, however, not be equated with a traditional 'lecturing and drill' approach in which the students remain passive. Active teachers ask a lot of questions (more than other teachers) and involve students in class discussion. In this way students are kept involved in the lesson and the teacher has the chance to monitor children's understanding of the concepts taught. Individual work is only assigned after the teacher has made sure children have grasped the material sufficiently to be ready for it. In general, effective teachers have been found to teach a concept, then ask questions to test children's understanding and, if the material did not seem well understood, to re-teach the concept, followed by more monitoring (Brophy, 1986; Brophy & Good, 1986). Overall, it is clear that effective teaching is not only active, but interactive as well.

The UK knowledge base is, in contrast to the American, a highly restricted one, although there is evidence of considerable contemporary policy interest in 'teaching' (Galton, 1995) and some promising new research avenues being explored, particularly in the fields of teacher's conceptual and subject

knowledge in areas such as mathematics (Askew *et al*., 1997) and the variation in teachers' behaviours within lessons (Creemers & Reynolds, 1996) and its effects on levels of pupil achievement.

Early research attempts in the UK to relate pupils' achievement gains to the broad educational philosophies and practices of teachers rated as 'progressive' or 'traditional' had of course generated rather little success (Bennett, 1976) and were widely criticised. Whilst 'progressive' teachers had lower gains, interestingly one teacher with what could be called 'structured, consistent progressivism' as a philosophy/practice generated the highest learning gain. In any case, the amount of variation in achievement explained by variation in teaching 'style' was small.

Later came the notable ORACLE study, which involved a 'process–product' orientation similar to that of the American teacher effectiveness material above and which found that the 'class enquirer' category of teachers who utilised a high proportion of whole-class teaching were more effective in delivering gains in mathematics and language, but that this did *not* apply in reading. Interestingly, this finding is also reported by Borich (1996) in his analysis of evidence about 'subject-specific' teaching behaviours in reading and in mathematics, where he notes the importance for reading of the following three factors.

Instructional activity. Spending time discussing, explaining and questioning to stimulate cognitive processes and promote learner responding.
Interactive technique. Using cues and questions that require every student to attempt a response during reading instruction.
Questioning. Posing thought-provoking questions during reading instruction that require the student to predict, question, summarise and clarify what has been said.

Borich (1996) rightly notes the importance of the existence of a 'blend' of types of teaching if reading gain is to take place: of structure and whole-class direct instruction on the one hand with an exploratory, interactive approach using classroom discussions and student ideas on the other.

The major British study of teacher effectiveness is, of course, the Junior School Project (JSP) of Mortimore *et al*. (1988), which reported the following factors as of importance in terms of school effectiveness across all outcome areas, showing again the power of the kind of 'blend' of factors that was noted as effective above:

Consistency among teachers. Continuity of staffing had positive effects but pupils also performed better when the approach to teaching was consistent.
Structured sessions. Children performed better when their school day was structured in some way. In effective schools students' work was proactively organised by the teacher, who ensured there was plenty for them to do yet

allowed them some freedom within the structure. Negative effects were noted when children were given unlimited responsibility for a long list of tasks.

Intellectually challenging teaching. Student progress was greater where teachers were stimulating and enthusiastic. The incidence of higher order questions and statements and teachers frequently making children use powers of problem solving was seen to be vital.

A work-centred environment. This was characterised by a high level of student industry, with children enjoying their work and being eager to start new tasks. The noise level was low and movement around the class was usually work-related and not excessive.

A limited focus within sessions. Children progressed when teachers devoted their energies to one particular subject area and sometimes two. Student progress was marred when three or more subjects were running concurrently in the classroom.

Maximum communication between teachers and students. Children performed better the more communication they had with their teacher about the content of their work. Most teachers devoted most of their time to individuals, so each child could expect only a small number of contacts a day. Teachers who used opportunities to talk to the whole class generated higher progress.

Record keeping. The value of monitoring student progress was important in the head's role, but it was also an important aspect of teachers' planning and assessment.

Parental involvement. Schools with an informal open-door policy which encouraged parents to get involved in reading at home, helping in the classroom and on educational visits tended to be more effective.

The factors at classroom level in the JSP that were significantly related to pupil gain in reading specifically were quite similar to what one would expect from the literatures noted so far, including the significant positive effects of 'use of a single reading scheme', 'time communicating with the whole class', 'time spent on higher order communication', 'time spent on non-work feedback' and 'the use of single curriculum activities'. Significant negative effects were seen for 'teacher time used to control classes', 'time spent supervising work' (presumably in groups or individual sessions), 'selective use of language textbooks', 'pupils having responsibility for managing their own work over long periods of time' and the 'proportion of activities devoted to mixed curriculum areas'.

The academic research outlined above is echoed in many of its findings by the characteristics of successful teaching shown by the on-going system of school inspections (OfSTED, 1996). From this 'professional' rather than 'research' literature, the successful teaching of literacy in general is argued to be shown by:

- early identification of what pupils already know about language and any difficulties they are experiencing, followed by targeted and positive support which teaches them about the system of written language and how to recognise and correct their own errors;
- making initial and continuing progress in reading and writing for all pupils a central objective of the school;
- involving parents in positive and practical ways through discussions at school and work with pupils at home;
- being based on a teaching programme which is thoroughly planned, with clear learning objectives and which provides direct teaching and careful assessment through to the end of Key Stage 2;
- capitalising on pupils' enthusiasm for communication to make reading and writing more enjoyable;
- teaching all aspects of literacy explicitly, directly and intensively in their own right and creating deliberate opportunities in the teaching of other subjects to extend experience and consolidate skills;
- a good understanding of techniques for beginning reading and writing, of how to select and combine them and how to judge their impact;
- using carefully sequenced whole-class, group and individual work to focus on strategies and skills, with the teacher combining instruction, demonstration, questioning and discussion, providing structure for subsequent tasks and giving help and constructive response;
- making use of systematic records of progress to monitor pupils' strengths and weaknesses, to intervene in a discriminating way and to plan the next stage of work;
- making good use of classroom assistants and volunteers, briefing them on how to work with pupils and to record what they do.

The successful teaching of *reading* in particular:

- equips pupils at the earliest stage to draw on the sources of knowledge needed when reading for meaning, including phonic knowledge (simple and complex sound–symbol relationships), graphic knowledge (patterns within words), word recognition (a sight vocabulary which includes common features of words), grammatical knowledge (checking for sense through the ways words are organised) and contextual information (meaning derived from the test as a whole);
- continues the direct teaching of reading techniques through both key stages, building systematically on the skills pupils have learnt earlier in, for example, tackling unfamiliar words;
- provides a range of reading material, usually based around a core reading programme, but substantially enriched with other good quality material, including information texts;
- stimulates and requires good library use;

- extends pupils' reading by focused work on challenging texts with the whole class or in groups;
- involves frequent opportunities for pupils to hear, read and discuss texts and to think about the content and the language used;
- gives time for productive individual reading at school and at home and opportunities for pupils to share their response with others.

School effectiveness

In addition to the material on teacher effectiveness contributed by research and by professional educational communities, there has emerged in the last two decades a voluminous international literature about the characteristics of the schools, as well as the classrooms, that 'add value' to childrens' achievement. The literature has generally been produced by utilising the same 'input/process/output' paradigm that characterised research in the American teaching effectiveness tradition we noted above, with reading scores often used as the intake and outcome variables (see review in Reynolds & Cuttance, 1992; Reynolds *et al*., 1996a).

One review (Levine & Lezotte, 1990) synthesised and summarised the extant American research on the characteristics of unusually effective schools as follows.

A productive school climate and culture

- An orderly environment.
- Faculty commitment to a shared and articulated mission focussed on achievement.
- A problem solving orientation.
- Faculty cohesion, collaboration, consensus, communication and collegiality.
- Faculty input into decision making.
- School-wide emphasis on recognising positive performance.

A focus on student acquisition of central learning skills

- Maximum availability and use of time for learning.
- An emphasis on mastery of central learning skills.

The appropriate monitoring of student progress
Practice-oriented staff development at the school site
Outstanding leadership

- Vigorous selection and replacement of teachers.
- 'Maverick' orientation and buffering.
- Frequent, personal monitoring of school activities and sense making.

- High expenditure of time and energy for school improvement actions.
- Support for teachers.
- Acquisition of resources.
- Superior instructional leadership.
- Availability and effective utilisation of instructional support personnel.

Salient parent involvement
Effective instructional arrangements and implementation

- Successful grouping and related organisational arrangements.
- Appropriate pacing alignments.
- Active/enriched learning.
- Effective teaching practices.
- An emphasis on higher order learning in assessing instructional outcomes.
- Co-ordination in curriculum and instruction.
- Easy availability of abundant, appropriate instructional materials.
- Stealing time for reading, language and mathematics.

Crucially, the interface between the school and the classroom or teacher level has also been explored (Teddlie, 1994), showing that school processes such as staff induction, proactive staff appointments, the removal of staff performing under 'floor level' expectations, the support for staff through relevant, targetted in-service training and frequent personal monitoring of and attention to the learning level by the principal and other senior staff are all school-level policies that can impact upon the learning level.

In the UK, school effectiveness factors generalised across the primary/secondary sectors and across curriculum areas such as English and mathematics are argued to be (Sammons *et al*., 1995; Reynolds *et al*., 1996b):

- headteacher leadership, goal setting and 'mission setting' combined with the involvement of staff;
- shared vision and goals amongst staff;
- a high quality learning environment;
- high quality teaching and learning;
- high expectations of children's possible achievements;
- the use of positive reinforcement and rewards;
- the careful monitoring of childrens' progress;
- attention to childrens' rights and responsibilities;
- purposeful teaching;
- high levels of parental involvement;
- high quality staff development.

For school effectiveness in the primary school sector the key study is again the JSP of Mortimore *et al*. (1988). In addition to the general, across-subject

factors noted above in our section on teaching effectiveness, significant school level policy factors associated with reading gain were:

- a headteacher who influenced teachers' record keeping;
- a headteacher who influenced teachers' teaching strategies;
- consistency between teachers;
- teacher involvement in decision taking;
- deputy headteacher involvement in decision taking;
- teachers having regular non-teaching lessons;
- a headteacher who encouraged teacher forward planning;
- a headteacher who influenced curriculum content.

Negative associations with reading gain were reported for:

- use of reading tests;
- variation between teachers in school guidelines;
- a headteacher who encouraged indiscriminate in-service course attendance.

A recent hitherto unpublished study conducted for the Literacy Task Force (McCallum, 1997) studies both the school effectiveness and teacher effectiveness 'levels' in a study of four primary schools with high Key Stage 2 English results and (in two cases) with teaching of literacy that had been commended by OfSTED in recent inspections. Data collected by classroom observation, interviews with school personnel, a visit to the school library and analysis of documentation revealed the following 13 factors found in all four (or in some cases three of the four) schools.

Literacy is given high status

- Emphasis on literacy in the stated aims of the school, in School Development Planning, in spending on resources and personnel.
- Specific timetabled slots for reading and writing skills in the morning.
- Classrooms and other areas of the school set up as 'language environments'.
- Well-developed libraries.
- Organised home reading schemes with leaflets to parents.

Headteachers have used staff deployment and pupil organisation to give the best possible chance for learning

- Deployment of staff to make best use of teachers' skills either with a particular class or to support other teachers; careful consideration of who teaches nursery and reception, who is the senior co-ordinator.
- Trained/well-briefed primary helpers assigned to one particular class.

- Management of pupil organisation to offer the best chance for learning to all pupils; flexible grouping and setting.
- Management of budget by the headteacher to ensure, as a priority, adequately staffed groups and smallish classes or groups.
- Visible presence of heads in classrooms, either teaching or observing.

There is a subject-based approach to the national curriculum
Elements of the English curriculum are taught separately

- Reading activities, spelling, punctuation and grammar, vocabulary and handwriting are separately identified in the teachers' timetables.

Schools have a culture of 'making things better'

- Heads and teachers 'care about school results'.
- Teachers analyse children's learning strategies.
- Heads analyse and track reading test results and act on findings.
- Early identification of children with special needs.
- Funds are targeted at support teachers for children with SEN, including able children.

Schools have a collaborative culture

- Staff feel supported by colleagues.
- Advice is sought and ideas adapted from senior co-ordinators, Section 11 teachers and reading recovery teachers.

There is a core of experienced primary teachers

- Reasonable stability of a core of staff keenly interested in literacy.

Teachers are united, committed and enthusiastic about literacy

- Teachers share the same philosophy: reading skills are taught using phonics and word recognition of 100 most common words; use of two main reading schemes plus a variety of supplementary material;
- Teachers prioritise the teaching of literacy skills.

Teaching is targetted, tightly planned, brisk, motivating and interactive

- Outstanding or very good teaching in three quarters of lessons observed.
- Clear targets set for all children on SEN register.
- Well-pitched tasks for the more able.
- Tasks are varied, achievable and interesting to children.
- Teachers are very active and constantly interact with children.

There is strong leadership and guidance in English

- Clear policies on each attainment target in English, showing progression by year group and giving detailed guidance for teachers.
- Clear policies on each element of writing showing progression by year group and giving detailed guidance for teachers.
- Clear library skills policies showing progression by year group and giving detailed guidance for teachers.
- Heads and co-ordinators offer strong leadership and teachers appreciate and follow their guidance.

Children's progression is monitored

- Monitoring and mentoring systems, often with non-contact time to review progression in other children's work, are in place.

Baseline testing is in place

- Reception children do a baseline test.

Reading and writing is regularly assessed and cumulative records are used

- Weekly spelling tests, regular teaching tests and regular teacher assessments in writing take place;
- SAT results and SAT papers are analysed and the findings used to inform teaching;
- Teachers trust each others' judgements.

From research to policy?

The level of agreement across the various studies on the effectiveness factors at the teaching level and at the school level outlined above is clearly considerable, as is the overlap of both these sets of factors with the teacher and school factors shown in the Literacy Task Force research. From all this it is clearly possible to argue that there exists a 'known to be valid' collection of methods which can be given to all schools to improve their English test results (as a surrogate for literacy). Indeed, the National Literacy Project represents a fusion in practice of many of these 'effectiveness' factors, with its daily literacy hour, its term-by-term planning of knowledge/skills for children of different ages and its teaching methodology of both whole-class teaching and structured group activities through differentiated ability groups.

However, there are a number of areas where researchers' findings and the present range of policies on literacy may sit uneasily together. Firstly, there is growing evidence of 'context specificity' in the precise factors associated

with learning gains, originally shown in interesting research from California, where highly effective schools in poor catchment areas pursued policies *discouraging* parental involvement in the school, in contrast to the effective schools in more advantaged catchment areas that encouraged the practice (Hallinger & Murphy, 1986). Whilst some factors apply across all social contexts (such as having high expectations of what children can achieve at 'school level' or 'lesson structure' at classroom level), it may be that certain factors apply only in certain environmental contexts. At classroom level an example might be that the factor of 'proceeding in small steps with consolidation if necessary' is important for *all* children who are learning to read for the first time in all contexts, whilst in the contexts inhabited by lower social class or lower attaining children it may be necessary to ensure high reading gain through the use of small 'steps' for teaching *all* knowledge and not just knowledge that is new.

Borich (1996) gives the following summary of teacher factors that may be necessary to achieve high achievement gains in literacy in classrooms in two different social settings, those of low socio-economic status and middle/high socio-economic status. Effective practices within low socio-economic status contexts involve the teacher behaviours of:

- generating a warm and supportive affect by letting children know help is available;
- getting a response, any response, before moving on to the next bit of new material;
- presenting material in small bits, with a chance to practice before moving on;
- showing how bits fit together before moving on;
- emphasising knowledge and applications before abstraction, putting the concrete first;
- giving immediate help (through use of peers perhaps);
- generating strong structure, ground-flow and well-planned transitions;
- the use of individually differentiated material;
- the use of the experiences of pupils.

Effective practices within middle socio-economic status contexts involve the teacher behaviours of:

- requiring extended reasoning;
- posing questions that require associations and generalisations;
- giving difficult material;
- the use of projects that require independent judgment, discovery, problem solving and the use of original information;
- encouraging learners to take responsibility for their own learning;
- very rich verbalising.

At school level, there are also hints of a wide range of possible contextual factors that may determine the precise nature of those factors which are needed to generate effectiveness:

- the socio-economic status of the catchment area (Hallinger & Murphy, 1986; Teddlie & Stringfield, 1993);
- the level of effectiveness of the school (Hopkins, 1996);
- the trajectory of effectiveness of the school (i.e. improving, static, declining) (Stoll & Fink, 1996);
- the region of the school (Reynolds, 1990);
- the urban/rural status of the school (Teddlie & Stringfield, 1993);
- the religiosity of the school (Coleman *et al*., 1981);
- the culture/history of the school (Stoll & Fink, 1996);
- the primary/secondary status of the school (Teddlie & Reynolds, 1998).

One must be intellectually honest and note that we do not yet know the extent of 'context specificity' as against 'universality' in the full range of teacher and school effectiveness factors (the tendency in many countries to sample only from socially disadvantaged areas has reflected strong social concern, but generated weak science). Also, at the level of what it may take to *improve* schools in different contexts we are still in the conjecturing stage (Hopkins & Reynolds, 1998).

However, it is clear that the danger in the present range of educational policies being 'rolled out' in the area of primary school children's literacy is that they are predominantly undifferentiated ones which are being introduced into very different local school contexts. Certainly, one could predict that awareness of the 'technology' of school and teaching effectiveness in literacy provided on in service days would be useful to all schools, but without attention to the 'start points' that schools are at, it may be that an undifferentiated 'roll out' will have the effect of merely maximising pre-existing differences between schools in their literacy competencies and literacy outcomes.

The second area where there must be doubts as to the wisdom of contemporary policies in the areas of literacy relates to the contemporary policy concern with the school level rather than the learning or classroom level. Of course, there are educational policies which are expected to impact directly upon teaching quality and methods, such as the recently changed requirements for courses of initial teacher training, and indeed the content within some of the distance learning material going to schools for use in the 1998–1999 academic year (such as that on 'word level work' and 'the literacy hour' particularly) is designed to impact upon classroom teaching.

However, the great majority of the policy 'levers' being pulled are at the school level, such as school development plans and target setting, and at local education authority level, such as LEA development plans. The

problems with the mostly 'school level' orientation of contemporary policy and contemporary educational discourse as judged against the literature are as follows:

- within-school variation by department within secondary school and by teacher within primary school is much greater than the variation between schools on their 'mean' levels of achievement or 'value added' effectiveness (Fitz-Gibbon, 1996);
- the effect of the classroom level in those multi-level analyses that have been undertaken, since the introduction of this technique in the mid 1980s, is probably three to four times greater than that of the school level (Creemers, 1994).

Simply, the most important determinant of children's literacy outcomes, the nature of their classroom experiences, is being targetted less than are their schools and their LEAs. It may be, though, that a classroom or 'learning level' orientation would be more productive of literacy gains for the following reasons.

- The departmental level in a secondary school or 'year' level in a primary school is closer to the classroom level than is the school level, opening up the possibility of generating greater change in classrooms.
- Whilst not every school is an effective school, every school has within itself some practice that is more effective than some other practice. Many schools will have within themselves practice that is absolutely effective across all schools. With a within-school, 'learning level' orientation every school can work on its own internal conditions;
- Focusing 'within' schools may be a way of permitting greater levels of competence to emerge at the school level, since it is possible that the absence of strategic thinking at school level in many parts of the educational system is related to the overload of pressures among headteachers, who are having referred to them problems which should be dealt with by the day-to-day operation of the middle management system of departmental heads, year heads, subject co-ordinators and the like.
- Within-school units of policy intervention, such as years or subjects, are smaller and therefore potentially more malleable than those at 'whole school' level.
- Teachers in general, and those teachers in less effective settings in particular (Reynolds, 1991, 1996; Stoll & Myers, 1997), may be more influenced by classroom level policies that are close to their focal concerns of teaching and the curriculum, rather than by the policies that are 'managerial' and orientated to the school level.
- The possibility of obtaining 'school level to school level' transfer of good practice, plus any possible transfer from LEAs in connection with their

role as monitors of school quality through their involvement in the approval of school's development plans, may be more difficult than the possibility of obtaining 'within school' transfer of practice.

Whilst it is clearly important to maximise both the school level factors and the learning level factors in their effectiveness, it is important to note that the most powerful intervention strategies within the area of literacy, Reading Recovery (Clay, 1993) and the 'Success for All' programme of Slavin (1996), have a pronounced focus upon pulling the lever of the 'instructional' level, as well as ensuring school level conditions are conducive to reading instruction. Indeed, in these programmes, which generate both the highest levels of achievement gain in reading ever seen in educational research *and* achievement gains that are (most unusually) higher amongst initially low scoring children, the school level is seen as merely setting the conditions for effective learning to take place at the classroom or instructional level. The school level is simply not given, in these programmes, the 'independent' source of variance explained or policy power that it holds within contemporary British educational policies.

The third area of concern about the nature of contemporary policies related to literacy is the difficulty of getting the 'effectiveness' knowledge into schools. It is now axiomatic amongst researchers and practitioners in the field of school improvement that there needs to be a degree of ownership by institutions and individuals of the process of school improvement in order for there to be take-up of knowledge and for the passage of new ideas to take place from the 'implementation' phase to the 'institutionalisation' stage (Fullan, 1991; Hopkins *et al*., 1994). Indeed, the entire international improvement movement arose out of a recognition that 'you cannot mandate what matters' and that the externally generated curriculum reform and curriculum materials of the 1960s and 1970s were not picked up by practitioners to a very marked degree (Reynolds, 1988).

Currently, though, knowledge is being delivered to schools *without* significant practitioner input in terms of choice of appropriate knowledge, consultation upon the phasing of the inputs and the organisation of the strategies to be introduced at school level. Similarly, the need to ensure the long term developmental capacities of schools to move 'beyond' the relatively conceptually and practically simple material they are being given, towards an ability to generate their own context-specific 'advanced' knowledge of effective practices and the like, is not currently being addressed.

The danger of any possible teacher reluctance to embrace the 'technologies' of effective teaching and effective schooling is of course that they may continue to operate at an intuitive rather than empirical-rational level in their day to day practice. Cato (1992) noted that the teachers of literacy in their study were in fact sometimes operating with something close to the 'mixed methods' or 'blend' model of effective teaching outlined above. The

great majority of teachers used phonics, reading schemes, whole-language and 'real books' methods, even though these methods came from diametrically opposed *philosophies* of education. Only 4% used 'real books' exclusively, although 28% relied on reading schemes. Only a fraction of teachers in this study never used phonics.

However, in their choice of methods to employ in different situations, with different classes and on different days, there was a tendency for teachers to be guided only by intuition, because of the absence of other, more rational constructs for choice. Likewise, OfSTED inspection evidence suggests great uncertainty over the methods that teachers think they should use and very variable quality in the implementation of any and all of them. It may be that the problems of literacy, particularly in Key Stage 2, may not necessarily only be to do with the *validity* of the methods being utilised in schools, but more the *reliability* of their implementation. If this is the case, then it is clear any absence of engagement by teachers in the educational process they are to be involved with from summer term 1998 may have damaging consequences on the reliability of implementation of 'good practice' in the field of literacy.

Conclusions

We have outlined in this paper some of the present range of educational policies that are concerned with literacy in primary schools and the knowledge bases that have clearly been influential in determining these policies, taken from the research and practice base of school effectiveness and teacher effectiveness. In a number of areas the tenative findings from the research literature suggest somewhat different emphases and policies to those being pursued, especially in the areas of context specificity, the importance of the learning level and the potential need for teacher 'ownership' of improvement policies.

These areas where there are potential conflicts between the implications of the research and the intended policies are clearly in need of further elaboration. The extent of 'universality' and 'context specificity' in the factors leading to literacy gain is clearly unknown at present, as are the ways of impacting upon the learning or classroom level that will enhance effectiveness most. The problem of ensuring practitioners partially 'own' the improvement of their methods and their schools, whilst at the same time ensuring that they routinely receive knowledge bases that they have not produced but which they clearly need is also a vexed one.

Much of the data that is needed to resolve these issues, such as context specificity, are, however, already collected or are shortly to be collected routinely as part of the monitoring and testing programme of the National Literacy Project. The 'roll out' of the National Literacy Strategy itself provides a chance to compare the effectiveness of various different strategies of

school and teacher development, given that there is likely to be variation in procedures according to the ideology and beliefs of the various consultants who will be working to implement the programme within schools in the approximately 150 different LEAs that will exist from April 1998. In a very real sense the 'experiment of nature' that is the National Literacy Strategy should, given the likely programmatic variation within it and given the inevitable variation in context it will interact with, both arbitrate on many unresolved issues and inform both its own future development and the bodies of knowledge on teacher effectiveness and school effectiveness that have helped to shape it historically.

References

Askew, M., Rhodes, V., Brown, M., William, D. & Johnson, D. (1997) *Effective Teachers of Numeracy* (London, King's College London School of Education).

Bennett, N. (1976) *Teaching Styles and Pupil Progress* (London, Open Books).

Borich, G. (1996) *Effective Teaching Methods*, 3 edn (New York, NY, Macmillan).

Brophy, J. (1986) Teaching and learning mathematics: where research should be going, *Journal for Research in Mathematics Education*, 17, pp. 323–346.

Brophy, J. & Good, T.L. (1986) Teacher behaviour and student achievement, in: M. C. Wittrock (Ed.) *Handbook of Research on Teaching* (New York, NY, Macmillan).

Cato, V. (1992) *The Teaching of Initial Literacy: how do teachers do it?* (Slough, NFER).

Clay, M. (1993) *Reading Recovery: a guidebook for teachers in training* (Auckland, New Zealand, Heinemann).

Coleman, J., Hoffer, T. & Kilgore, S. (1981) *Public and Private Schools* (Chicago, IL, University of Chicago).

Cooney, T.J. (1994) Research and teacher education: in search of common ground, *Journal for Research in Mathematics Education*, 25, pp. 608–636.

Creemers, B. (1994) *The Effective Classroom* (London, Cassell).

Creemers, B.P.M. & Reynolds, D. (1996) Issues and implications of international effectiveness research, *International Journal of Education Research*, 25, pp. 257–266.

Department for Education and Employment (1997a) *Excellence in Schools* (London, HMSO).

Department for Education and Employment (1997b) *The Implementation of the National Literacy Strategy* (London, DfEE).

Evertson, C.M., Anderson, C.W., Anderson, L.M. & Brophy, J.E. (1980) Relationships between classroom behaviors and student outcomes in Junior High mathematics and English classes, *American Educational Research Journal*, 17, pp. 43–60.

Fitz-Gibbon, C.T. (1996) *Monitoring Education: indicators, quality and effectiveness* (London, Cassell).

Fullan, M. (1991) *The New Meaning of Educational Change* (London, Cassell).

Galton, M. (1995) *Crisis in the Primary Classroom* (London, David Fulton Publishers).

Griffin, G.A. & Barnes, S. (1986) Using research findings to change school and classroom practice: results of an experimental study, *American Educational Research Journal*, 23, pp. 572–586.

Hallinger, P. & Murphy, J. (1986) The social context of effective schools, *American Journal of Education*, 94, pp. 328–355.

Hopkins, D. (1996) Towards a theory for school improvement, in: J. Gray, D. Reynolds & C. Fitz-Gibbon (Eds) *Merging Traditions: the future of research on school effectiveness and school improvement* (London, Cassell).

Hopkins, D. & Reynolds, D. (1998) Moving on and moving up: confronting the complexities of school improvement, *Educational Research and Evaluation*, in press.

Hopkins, D., Ainscow, M. & West, M. (1994) *School Improvement in an Era of Change* (London, Cassell).

Institute of Education (1997) *A Reading Revolution (The Preliminary Report of the Literacy Task force)* (London, Institute of Education).

Levine, D.U. & Lezotte, L.W. (1990) *Unusually Effective Schools: a review and analysis of research and practice* (Madison, WI, The National Center for Effective Schools Research and Development).

McCallum, B. (1997) A report on literacy in four schools, unpublished research report for the Literacy Task Force.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988) *School Matters: the junior years* (Salisbury, Open Books).

OfSTED (1996) *Successful Teaching of Literacy and Numeracy in Primary Schools: a starting point* (London, OfSTED).

Reynolds, D. (1988) British school improvement research: the contribution of qualitative studies, *International Journal of Qualitative Studies in Education*, 1(2), pp. 143–154.

Reynolds, D. (1990) The great Welsh education debate, 1980–1990, *History of Education*, 19(3), pp. 251–260.

Reynolds, D. (1991) Changing ineffective schools, in: M. Ainscow (Ed.) *Effective Schools for All* (London, David Fulton).

Reynolds, D. (1996) Turning around ineffective schools: some evidence and some speculations, in: J. Gray, D. Reynolds, C. Fitz-Gibbon & D. Jesson (Eds) *Merging Traditions: the future of research on school effectiveness and school improvement* (London, Cassell).

Reynolds, D. & Cuttance, P. (1992) *School Effectiveness: research, policy and practice* (London, Cassell).

Reynolds, D. & Farrell, S. (1996) *Worlds Apart?—a review of international studies of educational achievement involving England* (London, HMSO for OfSTED).

Reynolds, D., Creemers, B.P.M., Bird, J. & Farrell, S. (1994) School effectiveness—the need for an international perspective, in: D. Reynolds, B.P.M. Creemers, P.S., Nesselrodt, E. C. Schaffer, S. Stringfield & C. Teddlie (Eds) *Advances in School Effectiveness Research and Practice*, pp. 183–201 (Oxford, Pergamon Press).

Reynolds, D., Sammons, P., Stoll, I., Barber, M. & Hillman, J. (1996a) School effectiveness and school improvement in the United Kingdom, *School Effectiveness and Improvement*, 7(2), pp. 133–158.

Reynolds, D., Creemers, B.P.M., Hopkins, D., Stoll, L. & Bollen, R. (1996b) *Making Good Schools* (London, Routledge).

Robinson, P. (1997) *Literacy and Numeracy and Economic Performance* (London, London School of Economics, Centre for Economic Performance).

Sammons, P., Hillman, J. & Mortimore, P. (1995) *Key Characteristics of Effective Schools: a review of school effectiveness research* (London, OfSTED).

Sammons, P., Thomas, S. & Mortimore, P. (1997) *Forging Links: effective schools and effective departments* (London, Paul Chapman).

Scheerens, J. (1992) *Effective Schooling: research, theory and practice* (London, Cassell).

Slavin, R.E. (1996) *Education for All* (Lisse, Swets and Zeitlinger).

Stoll, L. & Fink, D. (1996) *Changing Our Schools* (Buckingham, Open University Press).

Stoll, L. & Myers, K. (1997) *No Quick Fixes: perspectives on schools in difficulty* (Lewes, Falmer Press).

Teddlie, C. (1994) The study of context in school effects research: history, methods, results and theoretical implications, in: D. Reynolds, B. Creemers, P. Nesselrodt, G. Schaffer, S. Stringfield & C. Teddlie (Eds) *Advances in School Effectiveness Research and Practice*, pp. 85–119 (Oxford, Pergamon Press).

Teddlie, C. & Reynolds, D. (1998) *The International Handbook of School Effectiveness Research* (Lewes, Falmer Press), in press.

Teddlie, C. & Stringfield, S. (1993) *Schools Make a Difference: lessons learned from a 10-year study of school effects* (New York, NY, Teachers College Press).

Walberg, H.J. (1986) Syntheses of research on teaching, in: M.C. Wittrock (Ed.) *Handbook of Research on Teaching* (New York, NY, Macmillan).

64

RHYME AND ALLITERATION, PHONEME DETECTION, AND LEARNING TO READ

P. E. Bryant, M. MacLean, L. L. Bradley and J. Crossland

In this article, 3 views of the relation between various forms of phonological awareness (detection of rhyme and alliteration and detection of phonemes) and children's reading were tested. These are (a) that the experience of learning to read leads to phoneme awareness and that neither of these is connected to awareness of rhyme, (b) that sensitivity to rhyme leads to awareness of phonemes, which in turn affects reading, and (c) that rhyme makes a direct contribution to reading that is independent of the connection between reading and phoneme awareness. The results from a longitudinal study that monitored the phonological awareness and progress in reading and spelling of 65 children from the ages of 4 years 7 months to 6 years 7 months produced strong support for a combination of the 2nd and 3rd models and none at all for the 1st model.

Two facts about children's phonological skills have been established beyond any doubt in recent research. First, there is a definite development in children's phonological skills (Lomax & McGee, 1987; Rozin & Gleitman, 1977). As children grow older their ability to make judgments about small phonological segments improves. From a very early age they are able to isolate and detect relatively large units such as syllables, and they can recognize rhymes (Knafle, 1973, 1974; Lenel & Cantor, 1981; MacLean, Bryant, & Bradley, 1987). Rhymes involve units that can be called *intrasyllabic* (Treiman, 1985, 1987), and in terms of size, they are usually somewhere between a syllable and a phoneme; to recognize that *cat* and *mat* rhyme, one must detect at some level the common two-phoneme segment *at*.

In contrast, children usually find phoneme detection tasks too difficult until they reach school age and begin to read. The two most commonly used

Source: *Developmental Psychology*, 1990, 26(3), 429–438.

tests of children's ability to detect phonemes are *phoneme tapping* and *phoneme deletion*. In phoneme tapping, the child has to tap out the number of phonemes in words spoken to him or her (three taps for *cat*, four for *silk*); in phoneme deletion, the child is asked to subtract a phoneme from a word ("What would *cat* sound like if you took away the first sound?"). Phoneme tapping has proved difficult for children up to the age of 5 years (Liberman, Shankweiler, Fischer, & Carter, 1974; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977). Phoneme deletion tasks are also difficult (Bruce, 1964) and have been conquered by prereaders only after considerable practice (Content, Kolinsky, Morais, & Bertelson, 1986).

The second significant discovery about phonological development is that there is a striking relation between children's phonological skills and their success in reading (Bryant & Bradley, 1985; Wagner & Torgeson, 1987). The better children are at detecting syllables (Mann & Liberman, 1984), rhymes (Bradley, 1988c; Bradley & Bryant, 1983; Ellis & Large, 1987; Lundberg, Olofsson, & Wall, 1980), or phonemes (Lundberg et al., 1980; Stanovich, Cunningham, & Cramer, 1984; Tunmer & Nesdale, 1985), the quicker and more successful will be their progress with reading. This relationship holds even when extraneous variables such as IQ, social class, and memory (Bradley & Bryant, 1985; MacLean et al., 1987) are controlled. Furthermore, properly controlled studies of children with reading difficulties have established that many—although by no means all—of them are strikingly insensitive to rhyme (Bradley, 1988b; Bradley & Bryant, 1978) and to letter–sound associations (Baddeley, Ellis, Miles, & Lewis, 1982; Frith & Snowling, 1983).

There is also some evidence that this relationship is specific to reading. Bradley and Bryant (1985) found that children's rhyming skills predict success in reading but not in mathematics. So far there has been no attempt to find out whether phoneme detection measures predict progress in reading but not in other educational skills.

It is easy to suggest plausible reasons for the connection between phonological skills and reading. One possible reason, which concerns phoneme detection, is that children have to be aware of phonemes in order to understand the alphabet, because alphabetic letters by and large represent the phonemes in words. There is evidence for an association between awareness of phonemes and learning the alphabet: People from China who have learned just a logographic script (Read, Zhang, Nie, & Ding, 1986) and people from Japan who have learned a logographic script together with a syllabary (Mann, 1986) are often rather poor at detecting or manipulating the constituent phonemes in a word.

Another possible reason for the connection between phonological skills and reading involves rhyme and spelling categories. Words that have sounds in common, such as rhyming words, often share spelling sequences as well in their written form (e.g., the *_ight* sequence in *light* and *sight*). Goswami (1986, 1988) has shown that even beginning readers are aware of the connection

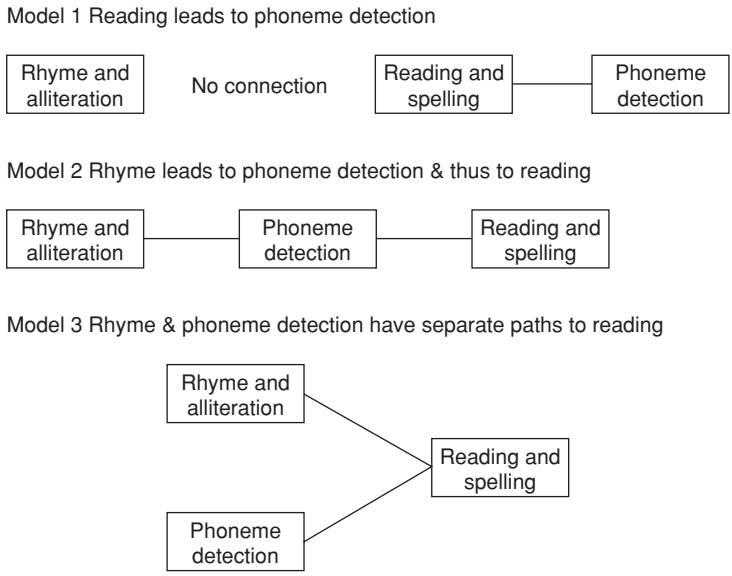


Figure 1 Three models of the links between phonological awareness and reading

between rhyme and spelling patterns. They make inferences about unfamiliar written words on the basis of rhyme: Learning to read *beak*, for example, helps them to read another new word like *peak*. Young children who are shown this connection make better progress in reading than do children who are not shown the connection (Bradley, 1988a).

The discovery of this strong and apparently specific connection is a notable success, but there has been much argument about the bearing early phonological development has on the relation between children's phonological skills and reading. Three different theories have been advanced that we refer to in Figure 1, as Models 1, 2, and 3.

Model 1 holds that the experience of being taught to read plays the main causal role. We owe this model mainly to the Brussels group (Morais, Alegria, & Content, 1987; Morais, Bertelson, Cary, & Alegria, 1986) who have argued that children acquire the ability to break up words into phonemes as a direct result of being taught to read. Model 1 states that awareness and segmentation of phonemes is the only relevant phonological skill as far as reading is concerned and that skills that younger children have, such as rhyme, are based on global perception rather than on analytic awareness and thus are too primitive to have an effect on reading. "Alphabetic literacy is (almost) a sufficient indication of segmental skill. . . . Rhyme appreciation and manipulation do not require segmental analysis" (Morais et al., 1987, p. 435).

Model 1 predicts that there should be no particular relation between the early skills (the detection of rhyme and alliteration) and the later ability to

detect phonemes because the two skills are unconnected and arise for different reasons. Rhyme develops naturally, whereas phoneme detection is the product of formal instruction. Model 1 also predicts that children's ability to detect phonemes should be far more strongly related to success in reading than should their rhyming skills. Indeed, Model 1 has some difficulty in dealing with the fact that measures of sensitivity to rhyme taken some time before children can read are highly successful predictors of their eventual progress in reading (Bradley & Bryant, 1983; Ellis & Large, 1987).

Model 2 gives a more important role to rhyme (Bryant & Bradley, 1985). According to Model 2, sensitivity to rhyme eventually leads to an awareness of phonemes, and this new skill in turn plays a role as the child learns to read and to spell. Thus, rhyme affects children's eventual success in reading, but it does so indirectly. Model 2 predicts a strong relation between children's early phonological skills, such as rhyme and alliteration, and later ones like phoneme detection, since the one set of skills leads to the other. Model 2 also predicts that children's early rhyme and alliteration scores should be related to their success in reading, but only because of the intervening development in phoneme detection. So the relation between rhyme and reading should disappear if controls are made for individual differences in phoneme detection.

In Model 3, rhyming affects reading directly. Model 3 follows Goswami's (1986, 1988) suggestion that children's sensitivity to rhyme makes a distinctive contribution to reading that is quite separate from the child's ability to isolate phonemes. Rhyme detection has a direct and distinctive effect by making children aware that words share segments of sounds (e.g., the *_ight* segment shared by *light*, *fight*, and *might*), and thus it prepares them for learning that such words often have spelling sequences in common too.

Model 3 produces one main prediction: a strong relation between children's early sensitivity to rhyme and their progress in reading, which will hold even after the effects of differences in the children's success in detecting phonemes have been controlled. Thus, the difference between Models 2 and 3 is that Model 2 predicts that controls for differences in the ability to detect phonemes will remove the relationship between rhyme and reading, whereas Model 3 holds that the relation will still be there after these controls.

The following is a longitudinal study in which these predictions were tested.

Method

Subjects

Ages

Of 66 children who took part in this project, we report data on 64. We failed to test one child on two tasks, and the other child left the country half way through the project. All but one of the children came from native English-

Table 1 Children's background measured by father's occupation and mother's education

Measure	*National percentage*	*Group's percentage*	*n*
Occupation			
Professional	5.6	3.1	2
Intermediate managerial	18.4	35.9	23
Nonmanual skilled	21.5	10.9	7
Manual skilled	31.1	32.8	21
Manual partly skilled	17.7	9.4	6
Unskilled	5.7	0.0	0
Single-mother families		7.9	5
Mother's education			
Degree(s)	7	18	12
Below degree level (HNC/Cert Ed.)	8	9	6
High school (age 18) A/ONC	5	12	8
High school (age 16) O/CSE	27	35	23
No qualifications	52	25	16

Note: HNC/Cert Ed. = technical or teacher's qualification; A/ONC = high school exam, 18 years; O/CSE = high school exam, 16 years.

speaking backgrounds. The exception is a boy whose mother is Swedish, and although English is the language spoken in his home, he knows some Swedish as well.

When the first phonological task that we report was given, the average age of the 64 children (33 girls and 31 boys) was 4 years 7 months (range = 4 years 2 months–5 years 3 months). We report data over a period of 2 years; when the last measure was taken, the average age of the 64 children was 6 years 7 months (range = 6 years 2 months–7 years 4 months). Earlier data from these subjects was reported in MacLean et al., 1987.

Social background

The children came from a wide range of backgrounds. Our measures of the home background included social class and the educational level of the parents (see Table 1). Intermediate and managerial occupations are overrepresented in our sample, and the children of unskilled men are underrepresented, given the national averages. However, the geographical area in which the research was carried out was a prosperous one, and so the sample was reasonably representative for the region.

We decided to use mothers' educational level as our measure of the children's background. We did not use social class because the project included several single-parent (mother) families to whom we could not apply the social-class index, which is based on the father's occupation. There were five different educational levels, but we could not treat them as a linear variable because we could not assume, for example, that the difference between Levels 1 and 2 was the same as that between Levels 2 and 3. Therefore, we treated this variable as a categorical one in our analyses.

I.Q. and vocabulary

When the children were 3 years 4 months old, we administered the British Picture Vocabulary Test (BPVS; a version of the Peabody Picture Vocabulary test standardized in Britain; Dunn & Dunn, 1982). The mean ratio score (average for the population is 100) on the BPVS was 104 (*SD* = 12.81). At 4 years 3 months, the children were given the full Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1963). The mean IQ was 110.94 (*SD* = 12.33). At 6 years 7 months (range = 6 years 2 months–7 years 1 months), the children were given the short version of the Wechsler Intelligence Scale for Children–Revised (WISC–R; Wechsler, 1974), either just before or just after the final session. The four WISC–R subtests given were Similarities, Vocabulary, Block Design, and Object Assembly. The mean prorated IQ was 111.84 (*SD* = 16.29). The relatively high IQ of the children reflects the relatively high proportion of middle-class children in the sample.

Procedure

The project was longitudinal; we report the results of four sessions when the children were 4 years 7 months, 5 years 7 months, 5 years 11 months, and 6 years 7 months. We had two sets of predictive measures and one set of outcome measures. The predictive measures were tests of rhyme and alliteration detection (given at ages 4 years 7 months and 5 years 7 months), and phoneme detection (at ages 5 years 7 months and 5 years 11 months). The outcome measures were reading, spelling, and arithmetic (at age 6 years 7 months). The first session was in the children's homes. At the second session, most children were seen at school. From then on, we saw everyone at school.

Detection of rhyme and alliteration

We gave the children versions of the rhyme-oddity task that had been used in previous studies (Bradley, 1980; Bradley & Bryant, 1983). The new feature of the version that we gave the children at 4 years 7 months used pictures to remove the memory load. The test consisted of 2 practice trials and then

10 experimental trials. In each trial, the child was given three words with pictures, two rhymed and the third did not (e.g., peg, cot, leg; fish, dish, book). The child's task was to tell us the one that did not rhyme. We also measured children's sensitivity to alliteration using the same methods. The children had to judge which of three words began with a different sound (e.g., pin, pig, tree; dog, sun, doll). The children's mean scores were 6.22 (SD = 2.63) out of 10 in the rhyme and 6.53 (SD = 2.44) in the alliteration-oddity test.

We devised a more difficult rhyme and alliteration task for the third session, when the children were 5 years 7 months, which involved explicit attention to the positions of sounds in words. We showed children three pictures (e.g., coat, coach, and boat) and asked them which one began with the same sound as, for example, *code* and ended with the same sound as *rote*. The children could not solve this task just on the basis of alliteration or rhyme because two words began like *code* and two rhymed with *rote*. The children were given two trials with feedback followed by seven without feedback. The reliability score for the seven nonfeedback trials proved to be lower than it was for all nine trials; therefore, we used the more reliable score. The mean score out of 9 was 4.88 (SD = 2.59; chance level = 3.0). The Spearman-Brown reliability coefficient for the test was .78.

Detection of phonemes

We used versions of the two most commonly used tests of phoneme detection —phoneme deletion and phoneme tapping tasks. These tasks did not involve a serious memory load because the child had to deal with only one word per trial.

For phoneme deletion at age 5 years 11 months, the children were introduced to a puppet that could not talk properly. They were told "She cannot say the beginnings of words, so if she wanted to say 'Hello, Ben' she would say 'ello, 'en.' Now you have the puppet. I'm going to say some words and you have to get the puppet to say them after me." The experimenter then gave four practice words with feedback, followed by five consonant, vowel, consonant (CVC) words and five CCVC words. This was the *first-sound* phoneme deletion task. In the *end-sound* phoneme deletion task, the children were told "Now the puppet can say the first sounds of words but cannot say the ends," and after some examples and practice trials with feedback they were given five CVC and five CVCC words. Words with blended consonants were too difficult, and we dealt only with the CVC scores in each task. The mean scores were 2.28 (SD = 2.15) out of 5 in the first-sound deletion test and 2.89 (SD = 1.79) in the end-sound deletion test. Random performance in such a task would not be far from zero. The Spearman–Brown reliability coefficients for these tests were, respectively, .94 and .76.

For phoneme tapping at 5 years 11 months, the children were given a stick to tap a block and were told that they would have to tap out the number of

sounds in words spoken to them. The experimenter gave several examples ("If I say *oo*, that has one sound and I tap once; if I say *boo*, I tap two times because that has two sounds; if I say *boot*, . . .) and then said "Now it's your turn. I'm going to say more sounds—you say them after me and tap at the same time." The sounds were one, two, or three phonemes in length. The mean score in this test was 7.36 (*SD* = 2.68) out of 12. The Spearman–Brown reliability coefficient for the test was .83.

None of the tests showed floor or ceiling effects. All of them produced reasonable reliability scores. Reliability was slightly higher in the phoneme detection tests than in the rhyme and alliteration and the joint rhyme/alliteration tests.

Reading, spelling, and arithmetic

In the final session, when the children were 6 years 7 months, we gave the following four tests:

1. France Primary Reading Test—a multiple-choice test with 48 items arranged in ascending difficulty to assess the understanding of words and simple sentences. The group's mean reading age on the test was 7 years 6 months (*SD* = 17.24 months).
2. Schonell Graded Word Reading Test plus extra words—a test involving reading single words. Because the test begins at a reading age of only 6 years, we added 10 words from a list of frequent words (Bradley, 1988a) and used the combined raw score. The group's mean reading age on the Schonell test was 7 years 2 months (*SD* = 15.27 months).
3. Schonell Spelling Test Form A plus extra words. We tested spelling with the Schonell Spelling Test, but preceded it with the same 10 additional frequent words. The group's mean spelling age on the Schonell test was 6 years 4 months (*SD* = 14.4 months).
4. WISC–R Arithmetic test. The mean scale score was 10.95 (*SD* = 2.99).

Results

Models 1, 2, and 3 illustrated in Figure 1 pose three questions: (a) Are rhyme detection and phoneme detection scores related? (b) Are these two sets of scores related to the children's success in reading and spelling? (c) Is there a connection between rhyme/alliteration and reading that is independent of children's ability to isolate single phonemes?

The relation between rhyme and phoneme detection

Model 1 predicts no relation, Model 2 predicts a strong relation, and Model 3 is neutral. Table 2 gives the correlations between phonological tasks and

Table 2 Correlations between phonological tasks

Task	*1*	*2*	*3*	*4*	*5*	*6*
1. Rhyme oddity (4 years 7 months)	—	.75	.69	.58	.33	.44
2. Alliteration oddity (4 years 7 months)		—	.67	.66	.52	.44
3. Joint rhyme/alliteration choice (5 years 7 months)			—	.65	.52	.61
4. Phoneme deletion, first sound (5 years 11 months)				—	.54	.60
5. Phoneme deletion, end sound (5 years 11 months)					—	.37
6. Phoneme tapping (5 years 11 months)						—

shows that the rhyme and alliteration measures are strongly related to the phoneme detection measures. Correlations, however, do not control for the effects of extraneous variables such as IQ or social class. For these controls we turned to multiple regressions. We report six five-step fixed-order multiple regressions. The dependent variable in each was one of the three measures of phoneme detection—first-sound and end-sound phoneme deletion and phoneme tapping—taken at 5 years 11 months.

The first four steps in each analysis were entered to control for differences in extraneous variables. These were (a) the children's ages, (b) their mothers' educational level, (c) their vocabulary (BPVS) and (d) their IQ (WPPSI). The fifth and final step was either the rhyme or alliteration oddity test given at 4 years 7 months or the joint rhyme/alliteration task given at 5 years 7 months. Thus, these analyses showed us whether the children's earlier rhyme and alliteration scores predicted their later skill at detecting phonemes, after the influence of differences in age, verbal skills, intelligence, and social background had been removed.

Table 3 shows that (a) the joint rhyme/alliteration task administered at 5 years 7 months is significantly related to all three phoneme detection measures and that (b) there is a strong connection between the rhyme oddity test given at 4 years 7 months and the first-sound phoneme deletion test given more than a year later. These significant connections are evidence that rhyme and alliteration are not, as Model 1 claims, separate from phoneme detection. There is a strong and—given the controls for IQ and social background—highly specific connection between the earlier rhyme and alliteration measures and the later tests of phoneme detection. The connections between rhyme and alliteration and the phoneme detection tasks (end-sound phoneme deletion and phoneme tapping [5 years 11 months] were less consistent, but

Table 3 Longitudinal prediction of performance in phoneme detection tasks by earlier rhyme and alliteration scores

	Phoneme deletion		
Dependent variable	*First sound*	*End sound*	*Phoneme tapping*
Steps 1–4 (R^2 change)			
Age	.01	.00	.07*
Mother's educational level	.40***	.28***	.26***
BPVS	.01	.00	.03
IQ[a]	.02	.02	.02
Step 5 (R^2 change)			
Rhyme oddity (4 years 7 months)	.13***	.02	.03
Alliteration oddity (4 years 7 months)	.09**	.05*	.02
Joint rhyme/alliteration choice (5 years 7 months)	.08**	.05*	.11**

Note: $df = 4$ for mothers' educational level; $df = 1$ for all other steps. BPVS = British Picture Vocabulary Scale.

[a] Wechsler Intelligence Scale for Children–Revised.

* $p < .05$. ** $p < .01$. *** $p < .001$.

the alliteration oddity test was significantly related to the end-sound phoneme deletion test. Thus, the regressions provide considerable support for Model 2 and virtually none for Model 1.

The relation of rhyme and phoneme detection scores to reading and spelling

Models 2 and 3 state that both types of detection task should predict reading, and because the hypothesis is about a specific connection, the models also predict that these detection tasks should be related only to reading and spelling, and not to arithmetic. On the other hand, Model 1 states that phoneme detection should be related to reading much more strongly than the prereading rhyme scores. Indeed, there is no reason for any connection between the rhyme scores and reading in Model 1.

Table 4 shows the correlations of the measures of rhyme and alliteration and phoneme detection with the children's ability to read, spell, and to do arithmetic at 6 years 7 months. The rhyme and alliteration oddity tasks and the joint rhyme/alliteration task are strongly related to all three reading/spelling measures.

Table 5 gives the results of 24 five-step fixed-order multiple regressions that tested whether the rhyme/alliteration and phoneme detection tasks predict

Table 4 Correlations between phonological measures and reading, spelling, and arithmetic

	Outcome measure (6 years 7 months)			
Predictive measure	*Schonell spelling*	*Schonell reading*	*France reading*	*Arithmetic*
Rhyme oddity (4 years 7 months)	.65	.64	.70	.53
Alliteration oddity (4 years 7 months)	.73	.79	.78	.48
Joint rhyme/alliteration choice (5 years 7 months)	.71	.72	.76	.58
Phoneme deletion, first sound (5 years 11 months)	.64	.67	.68	.46
Phoneme deletion, end sound (5 years 11 months)	.54	.58	.59	.33
Phoneme tapping (5 years 11 months)	.63	.61	.58	.54

performance in the reading, spelling, and arithmetic tests. The dependent variable in each regression was one of the four tests of reading, spelling, or arithmetic. The first four steps were the same as in the regressions reported in Table 3. The fifth step was one of the three rhyme/alliteration measures or one of the three phoneme detection measures. Thus, the regressions tested whether the phonological tasks predict reading, spelling, and arithmetic when vocabulary, intelligence, age, and social background are controlled.

Both the rhyme oddity task and the joint rhyme/alliteration task predict reading and spelling, but not arithmetic, and thus pass the test of specificity. However, the alliteration oddity task is not so specific a predictor of reading and spelling. It predicts reading and spelling particularly well ($p < .001$), but it is also related to arithmetic, although much less strongly ($p < .05$). It is worth noting how much of the variance in reading and spelling is accounted for in the three regressions in which alliteration was the final step: 71% of the variance in spelling, 76% in the Schonell test, and 74% in the France reading test.

All three phoneme detection tests were also significantly related to both reading measures. Two of them (first-sound phoneme deletion and phoneme tapping) were also significantly connected to spelling. However, the end-sound phoneme deletion test was not significantly related to spelling. Neither phoneme deletion test predicts arithmetic. So both pass the test of specificity, but the phoneme tapping test does not. It also predicts the children's arithmetic scores. There is a good reason why this test should also predict arithmetic. Tapping the right number of phonemes may depend on some

Table 5 Do phonological tests predict reading and spelling?

	Outcome measure (fixed order multiple regressions; 6 years 7 months)			
Dependent variable	*Schonell spelling*	*Schonell reading*	*France reading*	*Arithmetic*
	Steps 1–4 (R^2 change)			
Age	.04	.04	.00	—
Mother's educational level	.42**	.38**	.44***	.37**
BPVS	.00	.01	.02	.00
IQ	.14***	.17***	.13***	.06*
	Step 5 (R^2 change)			
Rhyme oddity (4 years 7 months)	.05**	.09***	.08***	.03
Alliteration oddity (4 years 7 months)	.11***	.17***	.15***	.05*
Joint rhyme/alliteration choice (5 years 7 months)	.07**	.09***	.10***	.02
Phoneme deletion, first sound (5 years 11 months)	.05**	.08***	.07**	.03
Phoneme deletion, end sound (5 years 11 months)	.01	.03*	.03*	.01
Phoneme tapping (5 years 11 months)	.05**	.04*	.03*	.08**

Note: BPVS = British Picture Vocabulary Scale.
* $p < .05$. ** $p < .01$. *** $p < .001$.

form of counting the phonemes; thus, the test may measure abilities related to number as well as to reading.

The results are a striking addition to the evidence of links between children's early phonological skills, particularly their sensitivity to rhyme and alliteration, and their eventual progress in reading.

The existence of a direct and independent connection between rhyme and alliteration and reading and spelling

Model 3 holds that rhyme makes a distinctive and direct contribution to learning how to read and that this contribution is quite independent of the children's sensitivity to phonemes. The fact that some rhyme and alliteration measures are better than the phoneme detection tasks at predicting reading and spelling already suggests that this direct link between rhyme/alliteration and reading does exist.

The most stringent test for a direct link would be fixed-order multiple regressions in which the dependent variable is a test of reading or spelling, the penultimate step is a phoneme detection test, and the final step is a test of rhyme or alliteration. This kind of analysis would show whether there is a connection between rhyme or alliteration and reading when differences in the ability to detect phonemes are controlled. If there is such a connection, it would be reasonable to conclude that part of the contribution that sensitivity to rhyme makes to reading has nothing to do with awareness of phonemes. It is worth noting that such a result could not be dismissed as a product of differences in reliabilities between rhyme and phoneme detection measures because the reliabilities of the measures in the penultimate step (the phoneme detection measures) were relatively high. Indeed, the reliability score for one of the phoneme detection measures—the first-sound phoneme deletion task—was .94 and was the highest among all of the phonological measures.

We carried out a series of six-step fixed-order multiple regressions with the following characteristics: (a) The dependent variable in each was one of the three reading and spelling measures, (b) the first four steps were the same as in the previous regressions, (c) the fifth and penultimate step was one of the three phoneme detection measures, and (d) the sixth and final step was one of the three rhyme and alliteration measures. That made 27 multiple regressions in all. Table 6 shows the variance due to the last two steps (the results for the first four steps are in Table 5) in these regressions.

Two points should be noted about the regressions. First, our measures account for an impressive amount of the variance in reading and spelling. In the regressions in which the Schonell reading and France reading tests were the dependent variables and the final step was the alliteration task given at 4 years 7 months, we account for 78% and 75%, respectively, of the variance in reading when the penultimate step was the first-sound phoneme deletion task.

Second, the rhyme and alliteration scores predict reading and spelling even after differences in phoneme detection are held constant. The two rhyme and alliteration scores are significantly related to France and Schonell reading performance in all 18 multiple regressions and, thus, whatever phoneme detection test was introduced as the penultimate test. Thus, rhyme and alliteration probably make a contribution to reading that is quite independent of the awareness of phonemes. The rhyme and alliteration scores were significant predictors in all but one of the six multiple regressions in which the dependent variable was spelling. As Table 6 shows, when the penultimate step was the first-sound phoneme deletion test, the rhyme oddity test no longer predicted spelling. However the joint rhyme/alliteration task did predict spelling in this case and when the other phoneme detection tests were entered as the penultimate step. We concluded that rhyme and alliteration definitely make an independent and distinctive contribution to reading and, almost certainly, to spelling as well.

Table 6 Relation of rhyme/alliteration to reading after controls for phoneme detection

	Schonell spelling		*Schonell reading*		*France reading*	
Variable	R^2 *change*	*Cumulative* R^2	R^2 *change*	*Cumulative* R^2	R^2 *change*	*Cumulative* R^2
Fifth step						
Phoneme deletion, first sound	.05**	.65	.08***	.67	.07***	.66
Sixth step						
Rhyme oddity	.02	.67	.04*	.71	.03*	.69
Alliteration oddity	.07**	.72	.11***	.78	.09***	.75
Joint rhyme/alliteration task	.03*	.68	.04*	.71	.05**	.71
Fifth step						
Phenome deletion, end sound	.01	.61	.03*	.62	.03*	.62
Sixth step						
Rhyme oddity	.04*	.65	.08***	.70	.07**	.69
Alliteration oddity	.10***	.71	.15***	.77	.13***	.75
Joint rhyme/alliteration task	.06**	.67	.07**	.69	.08***	.70
Fifth step						
Phoneme tapping	.05**	.65	.04*	.63	.03*	.62
Sixth step						
Rhyme oddity	.03*	.68	.07***	.70	.07**	.69
Alliteration oddity	.09***	.74	.15***	.78	.13***	.75
Joint rhyme/alliteration task	.03*	.68	.06**	.69	.07**	.69

* $p < .05$. ** $p < .01$. *** $p < .001$.

Combining Models 2 and 3—path analyses

The results have produced evidence for Model 2 (rhyme/alliteration scores are related to phoneme detection measurés, to reading and spelling but not to arithmetic) and Model 3 (rhyme/alliteration scores predict reading even after controls for differences in the ability to detect phonemes), but not for Model 1. It seems best to consider a combination of Models 2 and 3.

In the combined model, sensitivity to rhyme and alliteration makes two contributions to children's reading: (a) the indirect path—sensitivity to rhyme eventually leads to sensitivity to phonemes, which in turn helps the child learn about the alphabet and grapheme–phoneme correspondences—and (b) the direct path—sensitivity to rhyme makes a direct and distinctive contribution of the type originally suggested by Goswami (1986, 1988), which has

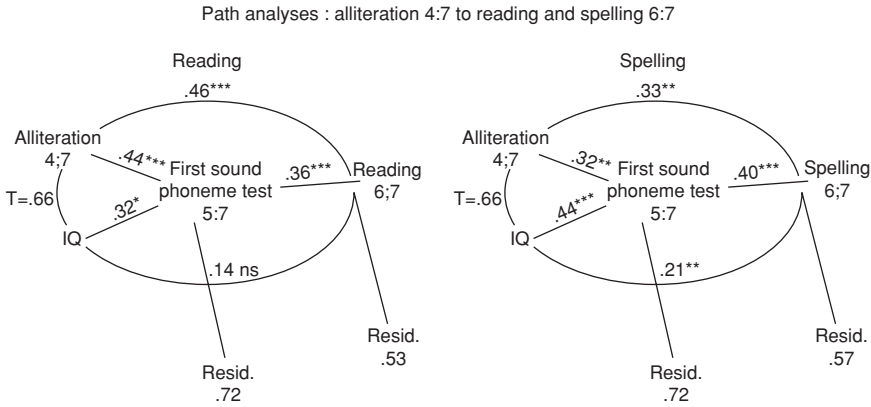


Figure 2 Path analyses of the connections between rhyme, phoneme detection, and reading. (Resid. = residual)

nothing to do with sensitivity to phonemes. We tested these models in a series of path coefficient analyses. These specific causal models were tested to examine the mediating effects of the phoneme detection tasks. The models were tested using the standardized scores of all measures in the analyses. Figure 2 gives two representative path analyses, in which the mediating variable of one is the joint rhyme/alliteration task and in the other is the phoneme deletion (first-sound) task. The path coefficients in the models are standardized partial regression coefficients (regression beta weights). The residual terms reflect the variance that is not explained in the model. All of the path analyses took this form: The final outcome measure was always one of the reading or spelling measures, and the mediating variable was either one of the phoneme detection tests or the joint rhyme/alliteration task. It can be seen in these two examples that both the indirect and direct paths were significant, as was the case in all of the path analyses.

Discussion

Our measures proved to be powerful predictors of reading and spelling. The multiple regressions, which included both a measure of rhyme or alliteration detection and one of phoneme detection, regularly accounted for above 65%, and in some cases for as much as 71%, of the variance in reading. So there certainly is a connection between early phonological skills and the child's progress in reading later on. We must now ask what form this connection takes? Although the mean IQ for the sample was above average, the children came from a wide range of family backgrounds; therefore, we can be reasonably sure that the connections that have been demonstrated are realistic and apply to the population at large.

The multiple regressions that accounted for so much of the variance also provided convincing evidence that rhyme and alliteration affects reading in two ways (and thus that both Models 2 and 3 are partially correct). There is a developmental path from early sensitivity to rhyme to awareness of phonemes a year or more later, and this awareness of phonemes is strongly related to reading. So the first route from early rhyming skills to reading is indirect and goes by way of phoneme detection.

Our analyses also demonstrate a direct connection between sensitivity to rhyme or alliteration and reading that appears to have nothing to do with phoneme detection. Our hypothesis is that the connection rests on the fact that words that rhyme or begin with the same sound, when written, often have spelling sequences of letters in common (e.g., *cat* and *hat* and, if one turns to beginning sounds, *peg* and *pen*). Of course, the connection between shared sounds and shared spelling patterns is not completely reliable. *Bite* and *light* rhyme, and *steak* and *beak* do not. Nevertheless, Goswami's (1986, 1988) research, mentioned earlier, shows that children use the fact that words with common sounds often have spelling sequences in common to help them to read or spell new words. There is, as well, a significant relationship between children's sensitivity to rhyme and their ability to make this type of inference about spelling patterns (Bradley, 1988c; Goswami & Bryant, 1988). It has been recognized for some time that children have to acquire other associations than single grapheme–phoneme correspondences and have to learn that associations with groups of letters probably form an important part of early reading (Gibson & Levin, 1976; Marsh & Desberg, 1983). It now seems that experiences with rhyme may play a key role in preparing children for this kind of learning.

We suggest that rhyme and alliteration affect reading at two phonological levels—first at the level of the phoneme, and second at the level that Treiman (1985, 1987) calls *intrasyllabic*. The combined model with its direct rhyme-to-reading path supports the suggestion that intrasyllabic units and their connection to whole spelling sequences like *_ight* play a crucial role in learning to read. There can be little doubt that this connection rests on children's awareness of rhyme and alliteration.

On the whole, the combined model and the results that led to it fit in well with previous research. The model gives a clear reason why in the past so many apparently disparate phonological measures have predicted reading successfully. It is even the case that parts of our combined model coincide with parts of the hypothesis advanced by the Brussels group (Morais et al. 1986, 1987), the proponents of Model 1. They too have argued and have produced convincing evidence for different levels of phonological skill, and they too have claimed that sensitivity to rhyme represents one level and awareness of phonemes quite another. However, our model disagrees with theirs on two points, both of which concern rhyme and alliteration. First, they have argued that sensitivity to rhyme and phoneme awareness have no

direct developmental link with each other because they have different origins. Rhyme, they think, develops naturally, whereas awareness of phonemes is a product of formal instruction and usually of instruction about reading. Our data, in contrast, suggest a strong developmental connection between the two types of phonological skill. Our second disagreement is about the connection between phonological skills and reading. The Brussels group have consistently argued that the only phonological achievement that can conceivably affect reading is the ability to detect and segment phonemes and thus to build up grapheme–phoneme correspondences. On the other hand, the data in our study confirm a strong connection between rhyme and reading and extend our knowledge about the influence of rhyme by demonstrating that some of that influence has nothing to do with single phonemes.

Our evidence does appear to differ from the results of some other research on one point. Although some studies have also shown strong connections between sensitivity to rhyme and reading, others have failed to do so. Lundberg (1987), for example, recently reported a correlation of only .22 between a rhyme test given to children of 6 years and a reading test which they took a year later. Stanovich, Cunningham, and Cramer (1984), as well, found little relation between a rhyme test and reading in a 5-year-old group, and their results appear to diverge from ours in another way: They did not find a relation between their rhyme test and some measures of phoneme detection. However, they pointed out that by 5 years of age the children tended to be at ceiling level in the rhyme test. This is also the probable explanation for Lundberg's result, inasmuch as his children were even a year older.

Work on the relation between performance in phonological tasks and success in reading has always rested on the assumption that the relation exists because children need to be able to break words down into constituent sounds when learning to read. If this is so, the relation should be specific to reading: Phonological tests should be related to reading but not to other aspects of education, such as arithmetic, which do not involve having to detect and manipulate the constituent sounds of words. It has already been shown that rhyming skills pass this specificity test: They do predict reading but not mathematics (Bradley & Bryant, 1985). Our study confirms that the relation between rhyme and reading is specific and shows as well that this specificity applies to the two phoneme deletion tests, as these also are related to reading and spelling but not to arithmetic. In stark contrast, the phoneme tapping test is strongly related to children's arithmetical skills as well as to their reading skills. This is probably because the task involves counting as well as the isolation of phonemes. This last result suggests that the tapping test is not as pure a test of phonemic awareness as has been suggested in the past (Tunmer & Nesdale, 1985).

Nevertheless, our study confirms the existence of a strong, consistent, and specific relation between children's phonological skills and reading. It also shows that rhyme and alliteration contribute to reading in at least two ways:

Sensitivity to rhyme and alliteration are developmental precursors of phoneme detection, which, in turn, plays a considerable role in learning to read. Sensitivity to rhyme also makes a direct contribution to reading, probably by helping children to group words with common spelling patterns. The study demonstrates the importance of early rhyming skills.

Acknowledgments

This research was supported by a grant from the Medical Research Council.

We are grateful for the help of Terezinha Carraher who read and commented on an earlier version of the article. We would like to thank the teachers and staff of local primary and first schools for letting us visit the children in our project at school.

References

Baddeley, A. D., Ellis, N. C., Miles, T. R., & Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition*, *11*, 185–199.

Bradley, L. (1980). *Assessing reading difficulties*. London: Macmillan Education.

Bradley, L. (1988a). Making connections in learning to read and to spell. *Applied Cognitive Psychology*, *2*, 3–18.

Bradley, L. (1988b). Predicting learning disability. In J. J. Dumont & H. Nakken (Eds.), *Learning disabilities: Vol. 2. Cognitive, social and remedial aspects* (pp. 1–17). Amsterdam: Swets.

Bradley, L. (1988c). Rhyme recognition and reading and spelling in young children. In R. L. Masland & M. R. Masland (Eds.), *Pre-school prevention of reading failure* (pp. 143–162). Parkton, MD: York Press.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature*, *271*, 746–747.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature*, *301*, 419–421.

Bradley, L., & Bryant, P. E. (1985). *Rhyme and reason in reading and spelling* (IARLD Monographs, No. 1). Ann Arbor: University of Michigan Press.

Bryant, P. E., & Bradley, L. (1985). *Children's reading problems*. Oxford, England: Blackwell's.

Bruce, D. J. (1964). The analysis of word sounds. *British Journal of Educational Psychology*, *34*, 158–170.

Content, A., Kolinsky, R., Morais, J., & Bertelson, P. (1986). Phonetic segmentation in pre-readers: Effect of corrective information. *Journal of Experimental Child Psychology*, *42*, 49–72.

Dunn, L. M., & Dunn, L. M. (1982). *British Picture Vocabulary Scale*. Slough, England: NFER-Nelson.

Ellis, N., & Large, B. (1987). The development of reading: As you seek so shall you find. *British Journal of Psychology*, *78*, 1–28.

Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology*, *1*, 329–342.

Gibson, E., & Levin, H. (1976). *The psychology of reading*. Cambridge, MA: MIT Press.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*, *42*, 73–83.

Goswami, U. (1988). Children's use of analogy in learning to spell. *British Journal of Developmental Psychology*, *6*, 21–34.

Goswami, U., & Bryant, P. E. (1988). Rhyming, analogy and children's reading. In P. B. Gough (Ed.), *Reading acquisition* (pp. 213–244). Hillsdale, NJ: Erlbaum.

Knafle, J. D. (1973). Auditory perception of rhyming in kindergarten children. *Journal of Speech and Hearing Research*, *16*, 482–487.

Knafle, J. D. (1974). Children's discrimination of rhyme. *Journal of Speech and Hearing Research*, *17*, 367–372.

Lenel, J. C., & Cantor, J. H. (1981). Rhyme recognition and phonemic perception in young children. *Journal of Psycholinguistic Research*, *10*, 57–68.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*, 201–12.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* (pp. 207–226). Hillsdale, NJ: Erlbaum.

Lomax, R. G., & McGee, L. M. (1987). Young children's concepts about print and reading: Toward a model of word reading acquisition. *Reading Research Quarterly*, *22*, 237–256.

Lundberg, I. (1987). Phonological awareness facilitates reading and spelling acquisition. In R. J. Bowler (Ed.). *Intimacy with language: A forgotten basic in teacher education* (pp. 56–63). Baltimore, MD: Orton Dyslexia Society.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology*, *21*, 159–173.

MacLean, M., Bryant, P. E., & Bradley, L. (1987). Rhymes, nursery rhymes and reading in early childhood. *Merrill-Palmer Quarterly*, *33*, 255–282.

Mann, V. (1986). Phonological awareness: The role of reading experience. *Cognition*, *24*, 65–92.

Mann, V., & Liberman, I. Y. (1984). Phonological awareness and verbal short term memory. *Journal of Learning Disabilities*, *17*, 592–599.

Marsh, G., & Desberg, P. (1983). The development of strategies in the acquisition of symbolic skills. In D. R. Rogers & J. A. Sloboda (Eds.), *The acquisition of symbolic skills* (pp. 149–154). New York: Plenum Press.

Morais, J., Alegria, J., & Content, A. (1987). The relationships between segmental analysis and alphabetic literacy: An interactive view. *Cahiers de Psychologie Cognitive*, *7*, 415–438.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*, 45–30.

Morais, J., Cary L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323–331.

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic spelling. *Cognition*, *24*, 31–34.

Rozin, P., & Gleitman, L. R. (1977). The structure and acquisition of reading: II. The reading process and the acquisition of the alphabetic principle. In A. S. Reber & D. L. Scarborough (Eds), *Toward a psychology of reading* (pp. 55–142). Hillsdale, NJ: Erlbaum.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. R. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology*, *38*, 175–190.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology*, *39*, 161–181.

Treiman, R. (1987). On the relationship between phonological awareness and literacy. *Cahiers de Psychologie Cognitive*, *7*, 524–529.

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology*, *77*, 417–427.

Wagner, R., & Torgeson, J. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, 192–212.

Wechsler, D. (1963). *Wechsler Preschool and Primary Scale of Intelligence*. New York: Psychological Corporation.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children–Revised*. Windsor, England: NFER.

65

WORD RECOGNITION

The interface of educational policies and scientific research

M. J. Adams and M. Bruck

As a result of a tremendous amount of research in educational, cognitive and developmental psychology on the nature and acquisition of reading skills, practitioners have a goldmine of evidence upon which to design effective educational programs for beginning and problem readers. This evidence is highly consistent in terms of delineating different stages of reading that young children pass through, the types of skills that they are to acquire, and the sorts of difficulties that they are likely to encounter. The purpose of this paper is to broadly outline current knowledge of the beginning stages of reading acquisition for both normal and problem readers and to relate this knowledge to current language arts curricular practices in North America.

Introduction

Across the centuries, methods to help the beginning reader attend to the sequences of letters and their correspondences to speech patterns have been a core element of most approaches to literacy instruction in alphabetic languages (Feitelson 1988; Mathews 1966; Richardson 1991; N. Smith 1974). We use the term 'phonics' to refer to such methods. In order to understand written text, the reader must be able to derive meaning from the strings of printed symbols on the page. Phonics methods are built on the recognition that the basic symbols – the graphemes – of alphabetic languages such as English encode phonological information. By making the relationships between spellings and sounds explicit, phonics methods are intended to assist the learning process by providing young readers and writers with a basis both for remembering the ordered identities of useful letter strings and for deriving the meanings of printed words that, though visually unfamiliar, are in their speaking and listening vocabularies.

Source: *Reading and Writing: An Interdisciplinary Journal*, 1993, 5, 113–139.

Despite their long and broad history, these traditional methods have periodically been challenged (see Balmuth 1982). The most recent attack comes from the Whole Language approach which has been adopted by many North American educational communities in the past decade or so. The Whole Language curriculum has several salient features that account for its popularity. First, it emphasizes teacher empowerment. Second, it advocates a child-centered method of instruction in which the child is seen as an active and thoughtful learner. Third, it stresses the importance of integrating reading and writing instruction, of drawing children quickly and clearly into the communicative and thought-worthy dimensions of print. In these ways, the Whole Language approach to literacy development tries to bridge the gap between written and oral language for the child by respecting her or his own intelligence, interest, and communicative competence.

While applauding the spirit of these goals, we also note that they are not, in themselves, incompatible with efforts to help children learn to understand and use the alphabetic principle. Rather the rift between Whole Language and phonics approaches derives from a deeper assumption of the movement. Specifically, Whole Language is anchored on the premise that there are strong parallels between reading acquisition and oral language acquisition. Goodman (1986), one of the fathers of this movement, specifically stresses the ease and naturalness of oral language acquisition and suggests that learning to read would be equally natural and simple if meaning and purpose were emphasized. By extension, it is argued, if reading for meaning is the very purpose of the exercise, then isn't it misguided, even counterproductive, to focus the reader's attention on the individual letters and their sounds? Viewing reading as a 'whole' integrated activity, major proponents of Whole Language decry the use of skill sequences and teaching skills in isolation. The more vocal advocates go so far as to claim that it is misguided to focus instruction on single words at all; because this will break up the text into meaningless pieces, their claim is that it will necessarily interfere with natural learning.

Frank Smith, another pioneer of the Whole Language movement, has asserted that 'decoding skills are used [by beginner readers] only to a very limited extent, and then primarily because a good deal of instructional effort is expended on impressing such methods on children' (1973: 71). Consistent with this, Smith also claims that the alphabetic principle is irrelevant to the fluent reader. Although he concedes that the mature reader may use decoding as a last resort to figure out unknown words, he argues that doing so is both rare and generally unnecessary. Instead, Smith suggests, skillful readers typically rely on the context and their knowledge of the world so as to gloss the words and guess the message. In this process, they are seen to sample a sparse minimum of graphic detail from the printed page – they do not visually process every word and they may not fully process any word. Instead they pick up only enough detail to corroborate or correct their hypotheses about the meaning and message of the text.

In fact, the basic conflict between the phonics and Whole Language approaches is also mirrored in the larger theoretical literature on reading. Across this century, a variety of models have been proposed that stipulate the cognitive processes that are necessary to create meaning from print. Generally, some models have placed the most relevant information and processes at the early stages of word recognition, whereas other models emphasize that the reader's understanding and perspective on meaning and message of the text are the crucial and basic elements of the reading process. The controversies and discrepancies among models concern the flow of information of the activities that occur during the entire process: Is the reader's understanding of the text generated bottom-up from print to meaning, or top-down from meaning to print? Are there alternate routes, short-cuts, or cognitive strategies for saving time and effort? More generally, what is involved and how it is learned? In this paper, we review three related bodies of evidence in order to evaluate the scientific assumptions and thus the validity of current practices in teaching young children to read. We discuss the most current models of reading, the importance of words and sublexical units in skilled reading, the parallels between oral and written language acquisition, the development of reading skills, and the basic difficulties associated with specific reading disabilities.

Recent models of reading

As a result of a convergence of a wide body of research along with significant advances in logical, mathematical, and computational sciences, recent models appear capable of mimicking the processes of reading and learning to read. These newer models, alternatively known as connectionist, neural net, or parallel distributed processing (PDP) models are built on the assumption that learning progresses as the learner comes to respond to the relationships between patterns or events. It is, for example, the overlearned relations among its edges that enables the infant to recognize a shape as a triangle, just as it is the overlearned relations among the letters of a printed string that enables the reader to recognize that string as a word. Similarly, it is the relations among the pitch, timing, and quality of its notes that evoke interest in a piece of music just as it is the relations among the meanings of its composite words that give texture and meaning to a sentence. (For a description of the logic and dynamics of these models, see Rumelhart & McClelland 1986; for an exploration of their pertinence to reading, see Adams 1990, and Seidenberg & McClelland 1989; for a discussion of their general importance and potential, see Bereiter 1991.)

In these models' portrayal of beginners or experts, the key is that they are neither top-down nor bottom-up in nature. Instead, all relevant processes are simultaneously active and interactive; all simultaneously issue and accommodate information to and from each other. The key to these models is not the

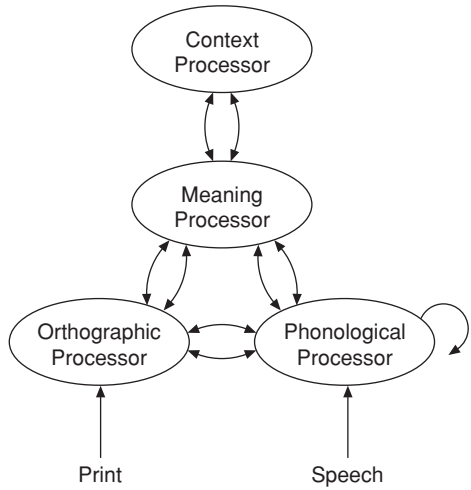


Figure 1 Schematic of the reading system (from Adams 1990: 158)

dominance of one form of knowledge over the others, but the coordination and cooperation of all with each other.

The architecture of one of these models of reading is schematized in Figure 1. Within each of the processors, knowledge is represented by many simpler units that have become linked or associated with one another through experience. The oval labeled Orthographic Processor, for example, represents the reader's knowledge of the visual images of word; with experience, individual letters are represented as interconnected bundles of more elementary visual features while printed words are represented as interconnected sets of letters. Similarly, the meanings of familiar word are represented in the Meaning Processor as bundles of simpler meaning elements while the pronunciation of words are represented as a complex of elementary speech sounds within the Phonological Processor.

Ultimately it is the links among clusters of one's knowledge – as they pass excitation and inhibition among each other – that are responsible for the fluency of the reader and the seeming coherence of the text. For the skillful reader, even as the letters of a fixated word are recognized, they activate the spelling patterns, pronunciations, and meanings with which they are compatible. At the same time, using its larger knowledge of language, life, and the text, the Context Processor swings its own bias among rival candidates so as to maintain the coherence of the message. Meanwhile, as each processor hones in on the word's identity, it relays its progress back to all of the other processors such that wherever hypotheses agree among processors, their resolution is speeded and strengthened. Guided by the connectivity both within and between processors, skillful readers are able to recognize

the spelling, sound, meaning, and contextual role of a familiar word almost automatically and simultaneously.

In many ways, this class of models fits well with Whole Language approaches. It emphatically asserts that literacy development depends critically and at every level on the child's interest and understanding of what is to be learned. Further, for learning to be efficient and productive, these models make clear that literacy cannot be fostered one piece at a time. The relations between the parts serve just as importantly in guiding the acquisition and refinement of the system as they do in its fluent operation. From the start, therefore, it is vital that literacy development involve reading, and writing, and spelling, and language play, and conceptual exploration, and all manner of engagement with text, in a relentlessly enlightened balance.

Indeed, although the value of many of the whole language initiatives are increasingly endorsed by both theory and research, there are exceptions. Most troublesome among these is the notion that reading is a 'psycholinguistic guessing game'. Again, according to this notion, readers pay little attention to the spellings of words, the spellings and spelling-sound correspondences of words are minimally relevant to learning to read, and given adequate exposure to meaningful engaging text, learning to read will proceed as naturally and autonomously as learning to talk. As we review below, in the 20 years since these ideas were first promulgated in Frank Smith's seminal book, *Understanding Reading* (1971), science has consistently, firmly, and indisputably, refuted these hypotheses (see Adams 1991, for a more detailed discussion of this saga).

The importance of words, spellings, and spelling-sound relations

Skillful reading, as it turns out, is scarcely a 'psycholinguistic guessing game', as Goodman (1967) termed it. Nor is it but incidentally visual as Frank Smith (1971) claimed. Instead, as is diagrammed in Figure 1, and physically must be the case, reading is visually driven. The letters and words of the text are the basic data of reading. For skillful adult readers, meaningful text, regardless of its ease or difficulty, is read through what is essentially a left to right, line by line, word by word process. In general, skillful readers visually process virtually each individual letter of every word they read, translating print to speech as they go. They do so whether they are reading isolated words or meaningful connected text. They do so regardless of the semantic, syntactic, or orthographic predictability of what they are reading (for reviews, see Just & Carpenter 1987, and Patterson & Coltheart 1987).

As these findings began to accumulate, researchers sought ways to dismiss them. Perhaps these findings reflected measurement error; perhaps they were misrepresentative, somehow brought on by one or another peculiarity of the laboratory tasks. Yet the findings that the skillful reader recognizes the words of a text on the basis of the sequences of individual letters that comprise

them were consistently replicated in a wide variety of paradigms in a number of laboratories. True, skillful readers neither look nor feel like they attend to the visual details of print as they read; but this, as it turns out, is the crowning explanation rather than the refutation of such findings. Readers must read the words just as listeners must hear them. It is only because readers (and listeners) process words so automatically and effortlessly that they have the mental time and capacity left to construct and reflect on that meaning and message. That is, the characteristic speed and effortlessness of skillful readers' word recognition is not simply a symptom or show of their skillful reading: It is necessary for its happening.

It is precisely through their words and wordings that speakers and authors strive to evoke and refine the meaning and message of their intentions. The words on the page are author's principle means of conveying their message: It will not do for readers to ignore them. Nor will guessing suffice: Even skillful adults are unable to guess correctly more than 25% of the time (Gough, Alford & Holley-Wilcox 1981). Furthermore, the process of guessing requires time and effort that can only be found at the expense of the normal processes of comprehension.

In fact, contrary to some of the common pronouncements of Whole Language mentors, skilled readers rely little on contextual cues to assist word identification. Rather, contextual cues contribute significantly to the speed and accuracy of word recognition only for those whose word identification skills are *poor* (e.g., Bruck 1990; Nicholson 1991; Perfetti, Goldman & Hogaboam 1979; Schwantes 1991; Stanovich 1981). This empirical finding is exactly opposite to that firmly espoused by educators such as Smith or Goodman who claim that poor readers' problems exist because they do not guess meanings of words from textual information. In fact just the opposite is true: It is the poorer less skilled reader who relies on contextual information to assist relatively poor word identification skills.

More specifically, it is skillful readers' overlearned knowledge about the sequences of letters and spelling patterns that enables them to process the print on a page so quickly and easily. As the reader fixates each word of text, the individual letters in focus are perceived almost instantly and effortlessly. Yet even as the letters are perceived, they are automatically clustered into familiar spelling patterns by virtue of the learned associations among them.

Such knowledge of spelling patterns is vital to the reader. It is responsible for protecting readers from misperceiving the order of letters within words (Adams 1981; Estes 1977) and for breaking long words into syllabic chunks even in the very course of perception (Mewhort & Campbell 1981; Seidenberg 1987). Similarly, it is this orthographic knowledge that causes skillful readers to look and feel as though they recognize frequent words holistically. Moreover, even where a word as a whole is not visually familiar, fragments of its spelling almost certainly will be. Referring back to Figure 1,

it is because of the overlearned connections between the Orthographic and Phonological Processors, that spelling patterns will all but automatically find their way to the Meaning Processor by way of their phonological equivalents so as to meet and thereby focus and strengthen the Orthographic Processor's tentative mapping of the word. In this way, the mature reader's deep and ready knowledge of spellings, sounds, and meanings virtually ensure that every foray into text will result in still more learning. In addition, it ensures that those many, many words of known meaning but incomplete visual familiarity may be read off with the ease and speed on which comprehension depends.

To construct understandings, the language comprehension system operates not on the meanings of individual words, but on the interrelations or overlap among them. Toward this end, comprehension works simultaneously with whole, cohesive grammatical units (whole phrases or sentences). Whether in listening or in reading, the process through which it does so is much the same (Jarvella 1971; Kleiman 1975). In either case, the words of the message are presented and perceived one by one. And although they are tentatively interpreted as they arrive, they are fully digested only after the clause or sentence is completely read or heard. In mystical deference to this process, speakers drop their pitch and pause at the end of every sentence: In this way, they let their listeners know that it is time to interpret while affording them time to do so. Mimicking this rhythm, skillful readers are found to march their eyes through all of the words of a sentence and then to pause at each period (Just & Carpenter 1987). It is during these end-of-sentence pauses that listeners or readers actively construct and reflect on their interpretations; it is during these interludes that they work out the collective meaning of the chain of words in memory and its contribution to their overall understanding of the conversation or text. Yet, in order for this interpretive process to succeed, the whole clause or sentence must still exist, more or less intact, in the listener's or reader's memory when she or he is ready to work on it. The quality of this representation is highly dependent upon the speed and effortlessness of the word recognition process. If it takes too long or too much effort for the reader to get from one end of the sentence to the other, the beginning will be lost from memory before the end has been registered.

This framework provides a powerful explanation for the findings of numerous studies that poor word identification skills are strongly coupled with poor reading comprehension in both children (Perfetti 1985; Rack, Snowling & Olson 1992; Stanovich 1982, 1991b; Vellutino 1991) and adults (Bruck 1990; Cunningham, Stanovich & Wilson 1990). In particular, it is not (as Smith and Goodman suggested) that skillful readers grasp the meaning of a text automatically and use it to figure out its words. Instead, they recognize the words automatically and use them to discern its meaning. In the end, the redundancy of text – of its syntax, semantics, and orthography – is highly functional not because it allows for skipping, but because it supplies the

superabundancy of information that protects the literal comprehension process from going astray.

Extending the analysis one step further, note that productive reading involves far more than literal comprehension. Rather, the priority issues while reading should include: Why am I reading this and how does this information relate to my reasons for so doing? What is the author's point of view, what are her or his underlying assumptions? Do I understand what the author is saying and why? Is the text internally consistent? Is it consistent with what I already know and believe or have learned elsewhere? If not, where does it depart and what can I think about the discrepancy? Comprehension in its truest sense is necessarily thought intensive. It requires analytic, evaluative, and reflective access to local and long-term memory. Yet, active attention is limited. To the extent that readers must struggle with the words, they necessarily lose track of meaning.

Is learning to read a natural biological process?

One of the major tenets of the Whole Language approach is that children are naturally predisposed to learn written language. The arguments are largely parallel to and, indeed, were spurred by those of Noam Chomsky (1965) that children are predisposed to learn spoken language. Chomsky's essential argument was that human language acquisition defied explanation through any simple model of learning. Human language was too rich and too varied, he argued. Whatever the units of learning might be, it was obviously impossible that language acquisition could be achieved through imitation, or by learning to connect units one-by-one. Furthermore, despite the complexity of the acquisition task, despite the noisiness and imperfection of the input to the child, despite the apparent absence of any universally endorsed instructional science on first-language acquisition, nearly all humans essentially master their native language within the first few years of life. (As Smith comments: 'There are relatively few books on such topics as *Why Johnny can't talk*' (1971: 49.))

The proposed solution was that babies were innately prepared to learn language. With a pre-wired 'Language Acquisition Device', human infants were seen to be endowed from birth with a deep knowledge of the essential physical, grammatical, and semantic components of all human languages. To become linguistically component in their native language, children need only discover which of the various options were operative in their own community of speakers. They did so, it was suggested, through a process of systematically testing, refining, and reformulating their built-in linguistic hypotheses (Chomsky 1965; McNeil 1970).

Frank Smith (1971) accepted these notions and imported them whole cloth to the literacy domain. Having already dismissed the utility of word- and letter-level instruction, the result was a broad disavowal of virtually

every sort of direct instruction. Children, Smith concluded, will best learn to read 'by experience in reading' . . . through ample, direct, and unmeditated engagement with meaningful text. Meanwhile, the teacher's most important job was one of providing feedback. But, he continued, it must be very sensitive feedback for, most of all, the teacher must create the sort of positive and supportive environment that would best encourage students to take on the risky business of testing new hypotheses (see Adams 1991).

Now, twenty years later, the notion that human babies are innately predisposed toward learning to speak has become generally accepted. However, it also seems that human grown-ups are naturally predisposed to help them out. Most parents tailor their speech to their babies' level. Perhaps unintentionally but both methodically and effectively, they do tutor their babies in the phonology, syntax, semantics, and pragmatics of their native language (e.g., Snow 1986).

Even so, with respect to literacy development, the most serious criticism of the let-them-learn-it-through-experience philosophy is that learning to read, unlike learning to talk, is not natural (see especially Liberman & Liberman 1992). Indeed, the parallelism between oral and written language acquisition that has been presumed by the Whole Language advocates must be seen as a flaw so serious as to undermine the whole approach. In Charles Perfetti's words:

> Learning to read is not like acquiring one's native language, no matter how much someone wishes it were so. Natural language is acquired quickly with a large biological contribution. Its forms are reinvented by every child exposed to a speech community in the first years of life. It is universal among human communities. By contrast literacy is a cultural invention. It is far from universal. And the biological contribution to the process has already been accounted for, once it is acknowledged that it depends on language rather than parallels it (1991: 75).

But if reading depends less on biological predispositions than on experience, then we are left with the question of what kinds of experience matter most. Within this question, we focus more directly on the issues of a phonics instruction and proceed by reviewing the literatures on normal development and on specific reading disabilities. In both of these discussions, we continue to refer to the model of reading proposed at the beginning of this paper.

Aspects of normal reading development

Beyond providing a coherent explanation of the nature of mature word recognition and its relation to the larger reading process, the connectionist framework carries several strong implications with respect to the nature of

the initial learning process. Again, the underlying assumption is that knowledge is encoded in the relations among the simpler aspects of one's experience. Understanding occurs as those relations are noticed; learning occurs as they are retained, strengthened, and enriched through repeated encounters and thought. By implication, understanding and learning rest on several basic prerequisites. First and most obviously, the student must be interested in what she or he is to learn. Beyond that, however, the student must also have a sense of which parts, elements, or aspects of the situation are relevant and of the kinds of interrelations among them that deserve attention.

Importantly, this business about knowing what to attend to is in no way unique to reading. You can watch a million football games and never get any better at following them unless you have had some sense of what to watch for along the way. Exposure alone is never sufficient. In addition, learners must somehow tune into the relations that carry and modulate information. Learning from and about written text depends on having a basic understanding of its forms, functions, and language. Learning to recognize printed words depends on noticing not just their meanings, but also their spellings, their sounds, and the relations between them. In this context, the critical lesson from research is that learning to read depends on certain insights and observations that, for many children, are simply not forthcoming without some special guidance.

Early reading development is often described in terms of a series of broad, overlapping stages (e.g., Chall 1983; Ehri 1992; Juel 1991; Gibson 1965; Gough & Hillinger 1980; Mason 1980) wherein the inception of each is marked by a qualitative change in the child's knowledge of how print works. While the fineness of the divisions between stages and even the foci of description within them differ from theorist to theorist, the child's discovery of the alphabetic principle is commony held to be a major milestone in the challenge of learning to read. In this section, we gloss the differences among theories in order to provide a broad overview of ways in which word recognition develops.

1. Fostering the emergence of early literacy knowledge

Before learning to read, most children develop insights as to the nature and functions of print. By reading books to children, they meet new creatures and characters and share their experiences. They learn new words, new language, and new concepts, and they also learn about the kinds of language, stories, and information that text can offer. They learn about decontextualized language, about the autonomy, authority, and permanence of the printed word, and they learn to create and comprehend realities beyond the here and now, realities that depend for their existence entirely on language (Snow & Ninio 1986). These sorts of understandings serve vitally to set up the knowledge, expectations, and interest on which learning to read depend. If children

also learn that reading is something they want to be able to do, they are well on their way.

Alongside this growing awareness of the nature and values of print, children also begin to learn how it works. They learn how to 'read' a book, which direction to hold it, which direction to turn the pages and which direction to read the words (e.g., Clay 1979; Downing 1979). They become aware of how print is formatted and that it encodes language. They become aware that its basic meaningful units are specific speakable words and that its words are comprised of letters. Importantly, all such learning is powerfully fostered by reading aloud to children, by engaging them regularly and interactively in the enjoyment and exploration of all manner of print (see Mason 1992). For example, simply providing little books for parent-child sharing has been shown to result in substantial increases in preschooler's knowledge of letters (McCormick & Mason 1986).

Preschoolers' familiarity with the letters of the alphabet is a powerful prognostic of the success with which they will learn to read (Bond & Dykstra 1967; Chall 1967). Beyond global correlations, young children's knowledge of letter names easily changes into interest in the sounds and in the spellings of words (Chomsky 1979; Mason 1980; Read 1971). In addition, knowing letters is strongly correlated with the ability to remember the forms of written words and the tendency to treat them as ordered sequences of letters rather than holistic patterns (Ehri 1992; Ehri & Wilce 1985). Conversely, not knowing letters is coupled with extreme difficulty in learning letter sounds (Mason 1980) and word recognition (Mason 1980; Sulzby 1983). Thus, finding ways to ensure that all children are developing a comfortable familiarity with letters should be a priority concern in all our preschools and kindergartens. There is, after all, no reason why playing with letters and print cannot be made as engaging and developmentally appropriate as sand tables, keys, and fruit salad.

In addition to supporting awareness of the forms, nature, and functions of print, exploration of text and language also helps preschool children to develop an awareness of the structure of their spoken language. They begin to understand the concept of a word, and that words and syllables are themselves made up of smaller sounds which can also be separated and rearranged (Bradley & Bryant 1983; Liberman, Shankweiler, Fischer & Carter 1974; Treiman 1985).

Because the printed symbols of alphabetic orthographies refer to phonemes, some have argued that awareness of phonemes is of particular importance for learning alphabetic orthographies (e.g., Liberman & Liberman 1992). It is the separable existence of the phonemes that seeds the connections from print to speech and that anchors the very logic of the writing system. In fact, faced with an alphabetic script, the child's level of phonemic awareness on entering school is widely held to be the strongest single predictor of the success she or he will experience in learning to read and of the likelihood that she or he will fail. This relationship has been demonstrated not only for

English (see, e.g., Blachman 1984; Juel 1991; Stanovich 1986), but also for Swedish (Lundberg, Olofsson & Wall 1980), Spanish (deManrique & Gramigna 1984), French (Alegria, Pignot & Morais 1982), Italian (Cossu, Shankweiler, Liberman, Tola & Katz 1988), and Russian (Elkonin 1973).

As it turns out, many of the activities (e.g., songs, chants, and word-sound games) that have long been enjoyed with preschoolers are ideally suited toward developing their sensitivity to the sound structure of language. Yet, all can be used with far more effectiveness if they are used with that goal in mind. By finger-pointing with print and by substituting and reordering words so as to turn sense to silliness, children can be led to discover the dependence of language on words. By exaggerating the meter of the songs and poems, children can be led to discover the existence of syllables. By contrasting rhyming words and playing with alliteration, they can be led to discover that syllables themselves can be teased apart. And having thereby introduced their essence, the phonemes can be more directly explored, separated, rearranged, and recombined. Kindergarten children who attend programs that emphasize such language play become significantly better readers and spellers when they move into the primary grades than children who are not offered such programs (e.g., Lundberg, Frost & Petersen 1988).

2. Helping young readers to break the code

In their initial efforts with print, many children rely solely on selective visual cues (Ehri 1992; Gough & Hillinger 1980), similar to those recommended by advocates of the 'psycho-linguistic guessing game'. Instead of examining words as a left-to-right sequence of letters, beginning readers tend to treat letter strings more as pictures (Byrne & Fielding-Barnesley 1989), basing recognition on the words' lengths, initial letters, or other distinctive features of their place or visual appearance. For the beginner with a limited reading vocabulary, the visual cue strategy might seem wholly serviceable. Yet, continued reliance on such partial visual cues eventually leads to severe difficulties in learning to read (Gough & Juel 1991; Snowling 1987).

Productive word learning in alphabetic orthographies ultimately depends on viewing words as a sequence of letters and associating their spellings with sounds. Some researchers believe that children first associate single phonemes with single graphemes, gradually learning to use orthographic units with experience (e.g., Bruck & Treiman 1992; Marsh, Desberg & Cooper 1977). Other researchers believe that even very beginning readers make associations between larger orthographic units, such as the rimes of words, and their sounds (e.g., Goswami & Bryant 1990). Despite these yet unresolved controversies surrounding the details of the process, it is important to emphasize that scientific research converges on the point that the association of spellings with sounds is a fundamental step in the early stages of literacy instruction. Furthermore, reading with fluency and comprehension depends not merely

on knowing about these relationships, but on using them, on overlearning, extending, and refining them, such that word recognition becomes fast and nearly effortless.

There are literally hundreds of articles to support these conclusions. Over and over, children's knowledge of the correspondences between spellings and sounds is found to predict the speed and accuracy with which they can read single words, while the speed and accuracy with which they can read single words is found to predict their ability to comprehend written text (see, e.g., Curtis 1980; Stanovich, Cunningham & Freeman 1984). Again, readers with fast and accurate word recognition skills have greater cognitive resources to direct attention to the meaning of text. Conversely, to the extent that children expend energy figuring out the identities of individual words, it can only be at the relative expense, in terms of time and mental capacity, of comprehending the meaning of the sentence or text.

For purposes of establishing the spelling-sound link, research indicates that teaching letters with sounds is more effective than teaching either alone (Ohnmacht 1969), that developing phonemic awareness in concert with letters and sounds is better than presenting letters and sounds alone (Ball & Blachman 1991), and that developing phonemic awareness with letters is more effective than developing phonemic awareness alone (Bradley & Bryant 1983; Byrne & Fielding-Barnesley 1991; Cunningham 1990). In short, the general pattern of results reasserts that development of skilled reading depends on the mastery of both the parts of the system and the functional relations among them. Again, these relations are just as important in guiding each other's acquisition as in supporting their fluent operation.

Significantly, research indicates that many of the same skills that underpin word recognition equally influence the acquisition of spelling skills (e.g., Waters, Bruck & Seidenberg 1985; Bruck & Treiman 1990; Griffith 1992). For example, Griffith (1992) found that phonemic awareness contributes directly and powerfully to beginners' ability to spell independently. And reciprocally, once children have established a basic awareness of phonemes and a willingness to print, independent writing is an excellent means of furthering both of these capacities. Moreover, because asking children to generate their own spellings is a way of engaging them in thinking actively and reflectively about the sounds of words in relation to their written representations, independent spelling can be an invaluable component of their phonics development.

Even so, independent spelling is not enough in itself. After all, not all the conventions of English orthography are intuitable. To read or write well, children must eventually learn how to spell correctly. At some point, therefor, they must be helped to do so. In developing children's spelling, one can alert them to such context-sensitive letter-sound rules as that /ē/ may be spelled with a *y* at the end of words but not at the beginnings. Further, structured spelling activities provide an ideal medium for analyzing and exploring those difficult consonant blends (e.g., rain, train, strain). Moreover, through

methodical use of word families, one can direct the children's attention to spelling patterns, ranging from the basics (e.g., pill, will, mill, . . . ; came, name, same . . .) to more subtle or sophisticated patterns (e.g., -tle, -ture, -tion).

Importantly, spelling instruction serves not only to improve the children's spelling but also to direct their attention to the range and composition of orthographic patterns that are to be perceptually consolidated in reading. In keeping with this, research indicates that more systematic attention to spelling results in exceptional progress in both reading and writing, especially for children who have already started independent writing (Uhry & Shepherd 1990).

As valuable as writing is, however, it is not enough. Young writers very often cannot read what they once wrote or even what they have just written (see, e.g., Chomsky 1979). In the end, reading with fluency and comprehension depends on a prodigious amount of perceptual learning. In significant measure, just as this learning is specific to reading, it can only be gained through reading.

The ability to decode proficiently and nondisruptively while reading depends integrally on familiarity, not just with individual letter-sound correspondences, but with the spelling patterns of which frequent words and syllables are comprised. Kindergartners and beginning first-graders are generally insensitive to the orthographic features of words as they tend to process all in a simple letter-by-letter manner (Bruck & Treiman 1992; Ehri & Robbins 1992; Juola, Schadler, Chabot & McCaughey 1978; Lefton & Spragins 1974; McCaughey, Juola, Schadler & Ward 1980). However, these children quickly show signs of their sensitivity to orthographic conventions. (Gibson, Osser & Pick 1963; Lefton & Spragins 1974) and to the frequencies with which spelling patterns occur within words (Treiman, Goswami & Bruck 1990). By third grade, normal readers exhibit adult patterns of responding with differential speed and ease to familiar words and typical spelling patterns (Backman, Bruck, Hebert & Seidenberg 1984).

From this perspective, phonics instruction per se takes on a very special value. In order to sound out a new word as they read, children must attend to each and every one of its letters, in left to right order. Each time they do so, the printed word will become more strongly and completely represented in memory so that, very soon, it will be recognized at a glance (see Ehri 1992). Of relevance, explicit, direct attention to phonics supports reading and spelling growth better than spelling instruction along with opportunistic attention to phonics while reading (Foorman, Francis, Novy & Liberman 1991).

Importantly, however, it is not just teaching children phonics that makes a difference but persuading them to use and extend it on their own, and a strong determinant of these tendencies lies in whether or not the children find it useful in their earliest efforts with print (Juel & Roper-Schneider 1985). More generally, there is strong rationale for ensuring that students'

first books consist largely of simple, short, and liberally repeated spelling patterns. To the extent that the new words in their texts are decodable, they reinforce the value as well as the process of approaching them as such. To the extent that these short simple patterns are basic, they will effectively anchor the longer, more complex, and less frequent patterns that are yet to be mastered.

3. Later stages

Relative to the overall literacy challenge, learning to recognize words is in fact a relatively small component. In terms of both linguistic and cognitive skills, there is so very much more to becoming a competent and productive reader and writer. In principle, because this is as true in speaking and listening as it is in reading or writing, one might expect the development of such knowledge to be reasonably independent of experience with text. In practice, however, there is a growing body of evidence to the contrary: Instead such knowledge appears strongly influenced by children's success with the initial hurdles of learning to read (for review and discussion, see especially Stanovich 1986, 1992).

Briefly the argument is as follows. Children who quickly master the early stages of reading find reading less aversive, less time-consuming, and more rewarding than those who do not. Because of this, better readers are likely to read more than children with poorer skills (Juel 1988) and, as a consequence, their early facility cascades into a sea of advantages. Most obviously, more reading is clearly the best path to better reading. In addition, however, through their experiences with text, these children acquire new language and vocabulary, new conceptual knowledge, new comprehension challenges, and new modes of thought to which they would not otherwise be exposed (e.g., Nagy, Anderson & Herman 1987). Meanwhile, to the extent that children struggle with reading, they can read far less even as they gain less from it. To the extent that they read less, even the opportunities are diminished.

Specific reading disabilities

Although some children seem to learn to read and write with remarkable ease, others have great difficulty. Population surveys suggest that between 7–15% of the school population suffers from specific reading disabilities, that is, from difficulties in learning to read and write that are not attributable to an identifiable intellectual deficiency, emotional disturbance, or other handicapping condition such as sensorial impairment or physical disability.

There are several important characteristics of these children that are important for the present discussion. The first is that although these children frequently come to our attention because they have difficulty understanding and producing written materials, their comprehension problems prove most

often to be symptoms of more basic word recognition difficulties (Backman et al. 1984; Perfetti 1985; Rack, Snowling & Olson 1992; Stanovich 1982; Vellutino 1979). These children often cannot understand text because they cannot read the words in text. Even when they accurately read or guess some of the words, their reading is so slow and belabored that the sense of the sentence or the larger logic of the passage is beyond reach; too much is forgotten along the way. Observations of such children provide compelling examples of how slow, inaccurate and effortful word recognition skills can impede comprehension.

Why do these children have so much trouble internalizing words? The position supported by the most empirical data is that these children have difficulties with the phonological aspects of word recognition. As evidenced through a number of different tasks and experimental paradigms, these children show exceptional difficulty in figuring out the correspondences between spelling and sounds (Olson 1985; Perfetti 1985; Snowling 1989; Vellutino 1979). At the same time, however, research attests that, like normal readers, disabled readers depend on spelling-sound correspondences for word recognition (e.g., Backman et al. 1984; Bruck 1988; Rack et al. 1992). A major difference between normal and disabled readers lies in the efficiency with which they construct, access, and learn these correspondences.

However, because the associations are so weak or incomplete, disabled children access this information in so highly inefficient a manner that all other components of the system suffer. According to the model depicted in Figure 1, dysfluencies with phonological processing inhibits the rapid flow of information back to the orthographic processor, which impedes the normal development and integration of orthographic information itself.

A second common manifestation of these children's phonological difficulties is poor awareness of the phonemic structure of language (Bradley & Bryant 1978; Bruck, in press; Bruck & Treiman 1990; Fox & Routh 1983; Mann & Liberman 1984; Snowling 1981). Indeed, this difficulty with phonemes is broadly held to underlie reading-disabled children's difficulty with spelling-sound relations. In keeping with this, programs that couple phonics with activities designed to enhance phonemic awareness are shown to be especially effective with disabled readers (Blachman 1987; Bradley & Bryant 1983; Williams 1980).

Distinct from phonological difficulties, research suggests that some disabled readers may be struggling against difficulties in visually registering or consolidating orthographic patterns (Bowers & Wolf 1992; Carr & Levy 1990; Reitsma 1989). In keeping with this, computer programs (which, one surmises, may have more potential for entrapping attention than paper exercises) designed to develop rapid, automatic responding to common spelling patterns are also found to result in significant improvements (Frederiksen, Warren & Roseberry 1985a, b; Roth & Beck 1987). Potential physiological basis for such difficulties has been forwarded by Lovegrove and his colleagues

(e.g., Lovegrove, Martin & Slaguis 1986; Williams & LaCluyse 1990). According to this position, reading-disabled children have problems in the early stages of visual processing which affect the quality and stability of the printed image (Breitmeyer 1989). Although Lovegrove's hypothesis is still too young to permit its sound evaluation, it is clear that problems at this early stage of processing would certainly and gravely affect all aspects of the reading process (e.g., Seidenberg 1992).

To be sure, the current literature weighs in more heavily on the side of the phonological than orthographic difficulties as the root cause of reading disabilities. Nevertheless, it is worth noting that the execution of many of the tasks that are used to assess phonological components per se (such as pseudoword reading) necessarily depend as well on orthographic knowledge and the relations between phonological and orthographic knowledge. Meanwhile, orthographic facility tends to be assessed through such tasks as recognition speed for previously viewed spellings and homonym discrimination; these are seemingly tasks that are clearly visual – except that visual learning about the precise order and identities of the letters in frequent words and spelling patterns is broadly held to be very strongly and perhaps inseparably dependent on sensitivity to the phonemic structures of words (see Adams 1990; Ehri 1992).

In any case, the connectionist framework described above makes clear that, whether or not it is possible to distinguish one of these root causes from another in assessment, deficits in either must inevitably and profoundly obstruct the operation and development of the system as a whole. Where orthographic images are incomplete, jumbled, or otherwise unstable, the establishment of firm, reliable links to the other processors will necessarily be impeded. Similarly, to the extent that phonological correspondences of spellings are slow or inaccessible, they cannot serve well to reinforce the spelling-to-word connections or, in the extreme, even to elicit the word in focus.

Within the connectionist framework, the escape from either impediment can be seen to lie in learning. That is, once the Orthographic Processor has thoroughly learned about the orthographic composition of the word or spelling pattern, that knowledge will serve to organize and beg completion of the print information as it arrives. Meanwhile, where orthographic learning is slow or incomplete, the connections between the Orthographic and Phonological Processor can provide critical support in stabilizing and reinforcing the image – provided that the child's knowledge of spelling-to-sound relations is developing properly. Similarly, for difficulties in accessing phonological information: Once the links between the Orthographic and the Phonological Processors have become firmly set up through learning, the phonological image of word or spelling pattern will be elicited automatically, no longer needing to be sought or constructed. But what about children who have difficulty accessing and analyzing the phonological image of a word? Without the ability to reflect easily on the phonological structure of the word,

how can such links be established in the first place? For these children, the framework suggests that the best and perhaps unique recourse may be had through the orthographic image and its piecewise links to the phonological processor. By understanding and studying the spelling-to-sound relations of the words they encounter, these children can refine and fill out their awareness of the words' phonological structure; meanwhile this very process serves equally to set the linkages that will hasten the words' recognition thereafter.

In keeping with the clinical data, the connectionist framework suggests that toward correcting weak or inappropriate learning in either the orthographic or phonological system, the necessary guidance and support are best, and uniquely, available by exercising its appropriate connections to the other. Nevertheless, it is worth noting that the model also carries significantly different prognoses for orthographic and phonological weaknesses. Children who fail to notice larger spelling patterns, concentrating instead on the decoding and synthesis of single letter-sound correspondences, will by virtue of sheer experience, sheer frequency of exposure, eventually learn the larger spelling patterns anyway. It's just that they will learn these patterns less quickly than if they had accorded them direct attention along the way; in addition, they may long fail to internalize less frequent spelling patterns. One might, for example, expect them to continue to pause on long and less frequent words and to commit phonetic but nonconventional spelling errors with less frequent patterns. In contrast, to the extent that children do not engage in spelling-to-sound translations but try to master their visual vocabularies through visual cues alone, they commensurately forfeit adequate opportunity for the spelling-to-sound connections to be established or refined. Yet, without the mnemonic support of the spelling-to-sound connections, the visual system must eventually become overwhelmed; the situation in which they are left is roughly analogous to learning 50,000 telephone numbers to the point of perfect recall and instant recognition.

Importantly, these are conjectures, speculations, based on the implications of the model and our understanding of the learning process. Even so, it is interesting to note their consistency with some recent work by Brian Byrne and his colleagues. In a cross-sectional study of second graders, Freebody & Byrne (1988) assessed second- and third-grade children's reading speed, reading comprehension, and phonemic awareness, along with their ability to read irregularly spelled words (assessing orthographic knowledge) and nonwords (assessing spelling-sound facility). From their performance on the latter two tasks, the children were divided into four groups: high on both, low on both, high on the irregular words but low on the nonwords (the 'Chinese'), and low on the irregular words but high on the nonwords (the 'Phoenicians'). Returning to the same children a year later, Byrne, Freebody & Gates (1992) found that while relative performance profiles of the first two groups had remained stable, those of the Chinese and the Phoenicians had shifted. As the Chinese children aged, their relative standing in terms of both word

recognition abilities and reading comprehension declined. In contrast, while the Phoenicians' reading speed remained relatively slow, they showed significant growth in both recognition abilities and in reading comprehension. Interestingly, both the Phoenicians and Chinese demonstrated adequate and, interestingly, *comparable* phonemic awareness – neither high like the generally good readers nor low like the generally poor readers. In view of this, Byrne et al. caution that phonemic awareness is necessary but not sufficient: Children must also learn to use spelling-sound correspondences in their efforts to read.

Note that, in contrast to this position, some have argued for approaches in which the treatment program is tailored to the perceptual strengths and styles of the learner. According to this view, phonics may be fine for children who are auditorily attuned and analytically natured. However, for those who are not so predisposed, then reading may better be developed without phonics, by emphasizing the global and visual dimensions of the challenge. Over the years, this argument has been broadly advocated and adopted. Even today its allure is saliently evidenced by the large following of Carbo, Dunn & Dunn (1986). Nevertheless, and despite the fact that many empirical studies have been conducted on this issue, there is little if any positive evidence for this sort of interaction between program effectiveness and preferred modalities (Arter & Jenkins 1977; Stahl 1988). Instead, reading-disabled children are commonly and repeatedly found to benefit most when given a reading program that directly emphasizes word recognition skills, rather than more general reading strategies (Lovett, Ransby & Barron 1988; Lovett, Ransby, Hardwick, Johns & Donaldson 1989).

A final important characteristic of the reading-disability syndrome concerns its persistence. Reading disability is not a condition that is specific to childhood nor one that disappears with development. Follow-up studies of reading-disabled children show that, despite all efforts, their problems may not disappear with age. Although many of these children become literate, it is typically with a great deal of effort. Furthermore, their word recognition deficits and associated phonological problems often persist into adulthood (Bruck 1985, 1990, in press; Labuda & DeFries 1988). Recognition of this fact forces one more extremely important question: To what extent could the prevalence or degree of reading disability be reduced by giving the children proper guidance early in the acquisition process? To what extent, for example, are the difficulties in remediation due to the fact that the children who come for help have already learned and overlearned other less efficient means of getting through text?

This question returns us to the topic of Whole Language and the disabled reader. To our knowledge, there have been no well-designed studies of how children with reading disabilities fare in Whole Language programs. For kindergartners, and especially for those who approach school with poorly developed concepts about print and language awareness, Whole Language programs may hold special value. In grade 1 and on, however, success depends

on breaking the code. Research strongly indicates that, without help, many children will not catch on to the alphabetic principle or its phonemic basis. Moreover, for those children who do not catch on, the disadvantages tend to spread themselves broadly and profoundly from reading to every other aspect of their schooling.

Whole Language without due attention to the code may place more children at risk for reading disabilities than would occur in more traditional programs. Wherever children who cannot discover the alphabetic principle independently are denied explicit instruction on the regularities and conventions of the letter strings, reading-disability may well be the eventual consequence. As yet, there are no data to confirm this hypothesis. Yet, they may soon be available if Whole Language adoptions continues to eliminate attention to phonics from the curriculum.

Finally, it is important to stress that reading-disabled students require much more than intensive phonics programs. They must also be given ample practice in reading and interpreting meaningful text. The model suggests and research firmly demonstrates that where the goal is to boost children's overall reading achievement, it is in fact best accomplished by engaging them with materials that are well beneath their frustration level: Regardless of how well a child reads already, high error rates are negatively correlated with growth; low error rates are positively correlated with growth (Rosenshine & Stevens 1984). Thus, while materials that are boring, uninformative, or otherwise inappropriate to the children's interests or dignity are to be avoided, neither can they be too hard. At the same time, however, we must find ways to support the children's linguistic, conceptual and cognitive growth beyond their reading levels. In view of this, most skilled clinicians also try to extend their reading-disabled students' linguistic, conceptual, and cognitive experience by providing tape-recorded books or giving parents materials to read to them.

Conclusions

Despite widespread and enthusiastic adoption of Whole Language approaches to reading instruction, there is surprisingly little scientific evidence to attest to its efficacy. As an exception, a synthesis by Stahl & Miller (1989) indicates the approach to have favorable outcomes when used in kindergarten and readiness programs. Given the Whole Language emphasis on the communicative functions and values of text, this finding is consistent with the literature at large: To learn to read, a child must learn first what it means to read and that she or he would like to be able to do so.

Indeed, at every level, leading children to explore the rewards and pleasures of text is of inestimable importance. But it is not enough: Research and theory have consistently and ever more emphatically affirmed and reaffirmed the importance of helping children to understand and to develop

ready working knowledge of the spellings and spelling-sound relations from which our writing system is built. Toward this end, kindergarten and preschool children benefit from activities designed to increase their awareness of the symbols and format of print and of the sound structure of spoken language. In first grade, independent reading and writing are the most valuable activities of all, but only if the child approaches them productively. In particular, research argues with indisputable strength that finding ways to induce young readers to attend to and to assimilate spellings and spelling-sound connections are of irreplaceable importance.

In fact, the Whole Language movement has brought with it many positive changes in classroom perspectives and practices. Why then couldn't its stand on phonics be revised so as to bring the best of both views to our classroom? In fact, this is broadly what is happening among good classroom teachers. But it is not a clear path.

There are several factors to account for the asymmetry between the research findings and some current educational practices involved in Whole Language. Many have noted that the Whole Language program is not simply an approach to teaching reading, rather it is a philosophical movement that addresses fundamental questions such as *What is reality? Where do facts come from? What is truth? How should power be distributed?* (Edelsky 1990: 7). These fundamental issues, it seems, promote an anti-research spirit within the Whole Language movement. Its leaders actively discredit traditional scientific research approaches to the study of development and more specifically to the evaluation of their programs. The movement's anti-scientific attitude forces research findings into the backroom making them socially and, thereby, intellectually unavailable to many educators who are involved in Whole Language programs. As a result, too many primary school teachers are now entering the field without fair education on how to teach or assess basic skills, much less on why or how they are important. This void is not only evident in the general curriculum for reading but also in the various programs suggested as remedial measures for reading disabled children. Keith Stanovich recently commented on this situation as follows:

> Sadly, very little of [the research of reading] had filtered through to reading teachers, parents and educational administrators. . . . It is also unfortunate that so little of this information has reached the somewhat separate groups of parents and special education personnel that deal with severe reading disability. Remedies for dyslexia are still more likely to emanate from cuckoo land than from the research literature (1991a: 79).

Yet, to the extent that this situation reduces to one of which of the adults win, too many children must lose. Reading disability threatens a child's entire education.

Over the last few decades, reading researchers have developed a far better understanding of the nature of print processing and how it feeds and fits into the rest of the reading system. They have learned why poor word recognition is a stumbling block for so many young readers and why, too, it is so frequently associated with poor comprehension. They have also learned much about how children learn to read words and how to help them do so. Educators can and should keep the positive initiatives of the Whole Language revolution. But is also time to put this knowledge about word recognition into college classrooms and into practice.

References

Adams, M. J. (1981). What good is orthographic redundancy? In: O. J. L. Tzeng & H. Singer (eds.), *Perception of print: Reading research in experimental psychology* (pp. 197–221). Hillsdale, NJ: Erlbaum Associates.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adams, M. (1991). Why not phonics *and* whole language? In: W. Ellis (ed.), *All language and the creation of literacy* (pp. 40–53). Baltimore, MD: The Orton Dyslexia Society.

Alegria, J., Pignot, E. & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. *Memory & Cognition* 10: 451–456.

Arter, J. A. & Jenkins, J. R. (1977). Examining the benefits and prevalence of modality considerations in special education. *Journal of Special Education* 11: 281–298.

Backman, J., Bruck, M., Hebert, M. & Seidenberg, M. (1984). Acquisition and use of spelling-sound correspondences in reading. *Journal of Experimental Child Psychology* 38: 114–133.

Ball, E. W. & Blachman, B. A. (1991). Does phoneme segmentation training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly* 26: 49–66.

Balmuth, M. (1982). *The roots of phonics*. New York: Teachers College Press.

Bereiter, C. (1991). Implications of connectionism for thinking about rules. *Educational Researcher* 20(3): 10–16.

Blachman, B. A. (1987). An alternative classroom reading program for learning disabled and other low-achieving children. In: W. Ellis (ed.), *Intimacy with language: A forgotten basic in teacher education* (pp. 271–287). Baltimore: Orton Dyslexia Society.

Blachman, B. A. (1984). Language analysis skills and early reading acquisition: In: G. Wallach & K. Butler (eds.), *Language learning disabilities in school-age children* (pp. 271–287). Baltimore: Williams and Wilkins.

Bond, G. L. & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly* 2: 5–142.

Bowers, P. G. & Wolf, M. (1992). Theoretical links between naming speed, precise timing mechanisms and orthographic skill in dyslexia (Submitted for publication).

Bradley, L. & Bryant, P. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature* 271: 746–747.

Bradley, L. & Bryant, P. (1983). Categorizing sounds and learning to read – A causal connection. *Nature* 310: 419–421.

Breitmeyer, B. (1989). A visually based deficit in specific reading disability. *The Irish Journal of Psychology* 10: 534–541.

Bruck, M. (1985). The adult functioning of children with specific learning disabilities. In: I. Sigel (ed.), *Advances in applied developmental psychology* (pp. 91–129). Norwood NJ: Ablex.

Bruck, M. (1988). The word recognition and spelling of dyslexic children. *Reading Research Quarterly* 23: 51–69.

Bruck, M. (1990). Word recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology* 26: 439–454.

Bruck, M. (in press). Persistent of dyslexics' phonological awareness deficits. *Developmental Psychology*.

Bruck, M. & Treiman, R. (1990). Phonological awareness and spelling in normal children and dyslexies: The case of initial consonant clusters. *Journal of Experimental Child Psychology* 50: 156–178.

Bruck, M. & Treiman, R. (1992). Learning to read: The limitations of analogies. *Reading Research Quarterly*.

Byrne, B. & Fielding-Barnesley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology* 83: 451–455.

Byrne, B. & Fielding-Barnesley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology* 81: 313–321.

Byrne, B., Freebody, P. & Gates, A. (1992). Longitudinal data on the relations of word-reaching strategies to comprehension, reading time, and phonemic awareness. *Reading Research Quarterly* 27: 140–151.

Carbo, M., Dunn, R. & Dunn, K. (1986). *Teaching students to read through their individual learning styles*. Englewood Cliffs, NJ: Prentice Hall.

Carr, T. H. & Levy, B. A. (1990). *Reading and its development*. Hillsdale NJ: Erlbaum Associates.

Chall, J. S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.

Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.

Chomsky, C. (1979). Approaching reading through invented spelling. In: L. B. Resnick & P. A. Weaver (eds.), *Theory and practice of early reading*, Vol. 2 (pp. 43–65). Hillsdale, NJ: Erlbaum Associates.

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Clay, M. (1979). *The early detection of reading difficulties*, 3rd ed. Portsmouth NH: Heinemann.

Cossu, G., Shankweiler, D., Liberman, I. Y., Tola, G. & Katz, L. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics* 9: 1–16.

Cunningham, A. E. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology* 50: 429–444.

Cunningham, A. E., Stanovich, K. E. & Wilson, M. (1990). Cognitive variation in adult college students differing in reading abilities. In: T. Carr & B. A. Levy (eds.), *Reading and its development: Component skills approaches* (pp. 129–159). San Diego: Academic Press.

Curtis, M. E. (1980). Development of components of reading skill. *Journal of Educational Psychology* 72: 656–669.

deManrique, A. M. B. & Gramigna, S. (1984). La segmentation fonologica y silabica en ninos de preescolar y primer grado. *Lectura y Vida* 5: 4–13.

Downing, J. (1979). *Reading and reasoning*. New York: Springer Verlag.

Edelsky, C. (1990). Whose agenda is this anyway? A response to McKenna, Robinson & Miller. *Educational Researcher* 19: 7–11.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In: P. B. Gough, L. C. Ehri & R. Treiman (eds.), *Reading acquisition* (pp. 107–144). Hillsdale, NJ: Erlbaum Associates.

Ehri, L. C. & Robbins, C. (1992). Beginners need some decoding skill to read words by analogy. *Reading Research Quarterly* 27: 12–26.

Ehri, L. C. & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly* 20: 163–179.

Elkonin, D. B. (1973). USSR. In: J. Downing (ed.), *Comparative Reading*. New York: Macmillan.

Estes, W. K. (1977). On the interaction of perception and memory in reading. In: D. LaBerge & S. J. Samuels (eds.), *Basic processes in reading* (pp. 1–25). Hillsdale, NJ: Erlbaum Associates.

Feitelson, D. (1988). *Facts and fads in beginning reading: A cross-language perspective*. Norwood, NJ: Ablex Publishing Corporation.

Foorman, B. R., Francis, D. J., Novy, D. M. & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology* 83: 456–469.

Fox, B. & Routh, D. (1983). Reading disability, phonemic analysis, and dysphonetic spelling: A follow-up study. *Journal of Clinical Child Psychology* 12: 28–32.

Frederiksen, J. R., Warren, B. M. & Rosebery, A. S. (1985a). A componential approach to training reading skills, Part 1: Perceptual units training *Cognition and Instruction* 2: 91–130.

Frederiksen, J. R., Warren, B. M. & Rosebery, A. S. (1985b). A componential approach to training reading skills, Part 2: Decoding and use of context. *Cognition and Instruction* 2: 271–338.

Freebody, P. & Byrne, B. (1988). Word reading strategies in elementary school children: Relationships to comprehension, reading time, and phonemic awareness. *Reading Research Quarterly* 24: 441–453.

Gibson, E. J. (1965). Learning to read. *Science* 148: 1066–1072.

Gibson, E. J., Osser, H. & Pick, A. D. (1963). A study of the development of grapheme-phoneme correspondences. *Journal of Verbal Learning and Verbal Behavior* 2: 142–146.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist* 4: 126–135.

Goodman, K. S. (1986). *What's whole in whole language?* Portsmouth, NH: Heinemann.

Goswami, U. & Bryant, P. (1990). *Phonological skills and learning to read*. Hillsdale, NJ: Erlbaum Associates.

Gough, P. B., Alford, J. A. & Holley-Wilcox, P. (1981). Words and contexts. In: O. J. L. Tzeng & H. Singer (eds.), *Perception of print: Reading research in experimental psychology* (pp. 85–102). Hillsdale, NJ: Erlbaum Associates.

Gough, P. B. & Hillinger, M. L. (1980). Learning to read: An unnatural act. *Bulletin of the Orton Society* 30: 179–196.

Gough, P. B. & Juel, C. (1991). The first stages of word recognition. In: L. Rieben & C. Perfetti (eds.), *Learning to read: Basic research and its implications* (pp. 47–56). Hillsdale, NJ: Erlbaum Associates.

Griffith, P. L. (1992). Phonemic awareness helps first graders invent spellings and third graders remember correct spellings. *Journal of Reading Behavior* 23: 215–234.

Jarvella, R. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior* 10: 409–416.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology* 80: 437–447.

Juel, C. (1991). Beginning reading. In: R. Barr, M. L. Kamil, P. B. Mosentha & P. D. Pearson (eds.), *Handbook of reading research*, Vol. 2 (pp. 759–788). New York: Longman.

Juel, C. & Roper-Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly* 20: 134–152.

Juola, J. F., Schadler, M., Chabot, R. J. & McCaughey, M. W. (1978). The development of visual information processing skills related to reading. *Journal of Experimental Child Psychology* 25: 459–476.

Just, M. A. & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.

Kleiman, G. M. (1975). Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior* 14: 323–339.

Labuda, M. & DeFries, J. C. (1988). Cognitive abilities in children with reading disabilities and controls: A follow-up study. *Journal of Learning Disabilities* 21: 562–566.

Lefton, L. A. & Spragins, A. B. (1974). Orthographic structure and reading experience affect the transfer from iconic to short-term memory. *Journal of Experimental Psychology* 103: 775–781.

Liberman, I. Y. & Liberman, A. M. (1992). Whole language vs. code emphasis: Underlying assumptions and their implications for reading instruction. In: P. B. Gough, L. C. Ehri & R. Treiman (eds.), *Reading acquisition* (pp. 343–366). Hillsdale, NJ: Erlbaum Associates.

Liberman, I. Y., Shankweiler, D., Fischer, F. W. & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology* 18: 201–202.

Lovegrove, W. J., Martin, R. & Slaghuis, W. (1986). A theoretical and experimental case for a visual deficit in specific reading disability. *Cognitive Neuropsychology* 3: 225–267.

Lovett, M., Ransby, M. J. & Barron, R. W. (1988). Treatment subtype and word type effects in dyslexic children's response to remediation. *Brain and Language* 34: 328–349.

Lovett, M. W., Ransby, M. J., Hardwick, N., Johns, M. S. & Donaldson, S. A. (1989). Can dyslexia be treated? Treatment specific and generalized treatment effects in dyslexic children's response to remediation. *Brain and Language* 37: 90–121.

Lundberg, I., Frost, J. & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly* 23: 263–284.

Lundberg, I., Olofsson, A. & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology* 21: 159–173.

McCaughey, M. W., Juola, J. F., Schadler, M. & Ward, N. J. (1980). Whole-word units are used before orthographic knowledge in perceptual development. *Journal of Experimental Child Psychology* 30: 411–421.

McCormick, C. E. & Mason, J. M. (1986). Intervention procedures for increasing preschool children's interest in and knowledge about reading. In: W. H. Teale & E. Sulzby (eds.), *Emergent literacy: Writing and reading* (pp. 90–115). Norwood, NJ: Ablex Publishing Corporation.

McNeil, D. (1970). The development of language. In: P. Mussen (ed.), *Carmichael's manual of child psychology*, Vol. 1 (pp. 1061–1162). New York: Wiley.

Mann, V. A. & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities* 17: 592–599.

Marsh, G., Desberg, P. & Cooper, J. (1977). Developmental strategies in reading. *Journal of Reading Behavior* 9: 391–394.

Mason, J. M. (1980). When do children begin to read: An exploration of our year old children's letter and word reading competencies. *Reading Research Quarterly* 15: 203–227.

Mason, J. M. (1992). Reading stories to preliterate children. In: P. B. Gough, L. C. Ehri & R. Treiman (eds.), *Reading acquisition* (pp. 215–241). Hillsdale, NJ: Erlbaum Associates.

Mathews, M. M. (1966). *Teaching to read: Historically considered*. Chicago: University of Chicago Press.

Mewhort, D. J. K. & Campbell, A. J. (1981). Toward a model of skilled reading: An analysis of performance in tachistoscoptic tasks. In: G. E. MacKinnon & T. G. Waller (eds.), *Reading research: Advances in theory and practice*, Vol. 3 (pp. 39–118). New York: Academic Press.

Nagy, W. E., Anderson, R. C. & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal* 24: 237–270.

Nicholson, T. (1991). Do children read words better in context or in lists? A classic study revisited. *Journal of Educational Psychology* 83: 444–450.

Ohnmacht, D. C. (1969). The effects of letter knowledge on achievement in reading in first grade. Paper presented at the American Educational Research Association, Los Angeles.

Olson, R. K. (1985). Disabled reading processes and cognitive profiles. In: D. Gray & J. Kavanagh (eds.), *Behavioral measures of dyslexia* (pp. 215–244). Parkton, MD:

Patterson, K. E. & Coltheart, V. (1987). Phonological processes in reading: A tutorial review. In: M. Coltheart (ed.), *Attention and performance*, Vol. 12: *The psychology of reading* (pp. 421–447). Hillsdale, NJ: Erlbaum Associates.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. (1991). The psychology, pedagogy, and politics of reading. *Psychological Sciences* 2: 70–76.

Perfetti, C. A., Goldman, S. & Hogaboam, T. (1979). Reading skill and the identification of words in discourse context. *Memory & Cognition* 77: 273–282.

Rack, J. P., Snowling, M. J. & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly* 27: 28–53.

Read, C. (1971). Preschool children's knowledge of English phonology. *Harvard Educational Review* 41: 1–34.

Reitsma, P. (1989). Orthographic memory and learning to read. In: P. G. Aaron & R. M. Joshi (eds.), *Reading and writing disorders in different orthographic systems* (pp. 51–73). Dordrecht/Boston/London: Kluwer Academic Publishers.

Richardson, S. O. (1991). Evolution of approaches to beginning reading and the need for diversification in education. In: W. Ellis (ed.), *All language and the creation of literacy* (pp. 1–8). Baltimore, MD: Orton Dyslexia Society.

Rosenshine, B. & Stevens, R. (1984). Classroom instruction in reading. In: P. D. Pearson, R. Barr, M. L. Kamil & P. Mosenthal (eds.), *Handbook of reading research* (pp. 745–799). New York: Longman.

Roth, S. F. & Beck, I. L. (1987). Theoretical and instructional implications of the assessment of two microcomputer word recognition programs. *Reading Research Quarterly* 22: 197–218.

Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tenses of English verbs. In: J. L. McClelland & D. E. Rumelhart (eds.), *Parallel distributed processing*, Vol. 2: *Psychological and biological models* (pp. 216–271). Cambridge MA: MIT Press.

Schwantes, F. M. (1991). Children's use of semantic and syntactic information for word recognition and determination of sentence meaningfulness. *Journal of Reading Behavior* 23: 335–350.

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy. In: M. Coltheart (ed.), *Attention & performance*, Vol. 12: *The psychology of reading* (pp. 245–263). Hillsdale, NJ: Erlbaum Associates.

Seidenberg, M. S. (1992). Dyslexia in a computational model of word recognition in reading. In: P. B. Gough, L. C. Ehri & R. Treiman (eds.), *Reading acquisition* (pp. 243–274). Hillsdale NJ: Erlbaum Associates.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review* 96: 523–568.

Smith, F. (1971). *Understanding reading.* New York: Holt, Rinehart & Winston.

Smith, F. (1973). *Psycholinguistics and reading*. New York: Holt, Rinehart & Winston.

Smith, N. B. (1974). *American reading instruction*. Newark, DE: International Reading Association.

Snow, C. E. (1986). Conversations with children. In: P. Fletcher & M. Garman (eds.), *Language Acquisition*, 2nd ed. (pp. 69–89). New York: Cambridge University Press.

Snow, C. E. & Ninio, A. (1986). The contracts of literacy: What children learn from learning to read books. In: W. H. Teale & E. Sulzby (eds.), *Emergent literacy: Writing and reading* (pp. 116–138). Norwood, NJ: Ablex Publishing Corporation.

Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research* 43: 219–234.

Snowling, M. J. (1987). *Dyslexia: A cognitive developmental perspective*. Oxford: Basil Blackwell.

Snowling, M. J. (1989). Developmental dyslexia: A cognitive developmental perspective. In: P. G. Aaron & R. M. Joshi (eds.), *Reading and writing disorders in different orthographic systems* (pp. 1–23). Dordrecht/Boston/London: Kluwer Academic Publishers.

Stahl, S. A. (1988). Is there evidence to support matching reading styles and initial reading methods? A reply to Carbo. *Phi Delta Kappan* 70: 317–322.

Stahl, S. A. & Miller, P. (1989). Whole language and language experience approaches for beginning reading: A quantitative research synthesis. *Review of Educational Research* 59: 87–116.

Stanovich, K. E. (1981). Attentional and automatic context effects in reading. In: A. Lesgold & C. Perfetti (eds.), *Interactive processes in reading* (pp. 241–263). Hillsdale, NJ: Erlbaum Associates.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading, Part 1: Word decoding. *Journal of Learning Disabilities* 15: 485–493.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in acquisition of literacy. *Reading Research Quarterly* 21: 360–407.

Stanovich, K. E. (1991a). Commentary: Cognitive science meets beginning reading. *Psychological Sciences* 2: 77–81.

Stanovich, K. E. (1991b). Word recognition: Changing perspectives. In: R. Barr, M. L. Kamil, P. Mosenthal & P. D. Pearson (eds.), *Handbook of reading research*, Vol. 2 (pp. 418–452). New York: Longman.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In: P. B. Gough, L. C. Ehri & R. Treiman (eds.), *Reading acquisition* (pp. 307–342). Hillsdale, NJ: Erlbaum Associates.

Stanovich, K. E., Cunningham, A. E. & Freeman, D. J. (1984). Relation between early reading acquisition and word decoding with and without context: A longitudinal study of first-grade children. *Journal of Educational Psychology* 76: 688–677.

Sulzby, E. (1983). A commentary on Ehri's critique of five studies related to letter-name knowledge and learning to read: Broadening the question. In: L. M. Gentile, M. L. Kamil & J. S. Blanchard (eds.), *Reading research revisited*. Columbus, OH: Charles E. Merrill.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology* 39: 161–181.

Treiman, R., Goswami, U. & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition* 18: 559–567.

Uhry, J. K. & Shepherd, M. J. (1990). The effect of segmentation/spelling training on the acquisition of beginning reading strategies. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Vellutino, F. (1979). *Dyslexia: Theory and research*. Cambridge MA: MIT Press.

Vellutino, F. R. (1991). Introduction to three studies on reading acquisition: Convergent findings on theoretical foundations of code-oriented versus whole-language approaches to reading instruction. *Journal of Educational Psychology* 83: 437–443.

Waters, G. S., Bruck, M. & Seidenberg, M. S. (1985). Do children use similar processes to read and spell words? *Journal of Experimental Child Psychology* 39: 511–530.

Williams, J. P. (1980). Teaching decoding with a special emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology* 72: 1–15.

Williams, M. & LeCluyse, K. (1990). Perceptual consequences of a temporal processing deficit in reading disabled children. *Journal of the American Optometric Society* 61: 111–121.

66

UNDERSTANDING OF CAUSAL EXPRESSIONS IN SKILLED AND LESS SKILLED TEXT COMPREHENDERS

J. Oakhill, N. Yuill and M. L. Donaldson

This experiment investigated the relation between 7- to 8-year-old children's reading comprehension and their understanding of causal expressions. A group of skilled comprehenders was compared to a less skilled group on two oral tasks involving *because* sentences: a questions task and a sentence completion task. For each task, the subjects received deductive items (where *because* introduces the evidence for a conclusion) and empirical items (where *because* introduces a cause). The experiment also investigated the effect of modifying the instructions for the deductive items so as to focus the subjects' attention on the source of evidence for a conclusion. Skilled comprehenders performed significantly better than less skilled comprehenders on the deductive, but not on the empirical, items. Performance on deductive items was poorer than that on empirical items. However, scores on the deductive items were increased by the modified instructions.

Understanding of causal expressions is obviously a basic and essential aspect of text comprehension. Unless connectives such as *because* are taken as signalling a causal link between the events in separate clauses, many of the text-connecting inferential links in the text will not be made by the reader.

A number of our previous studies have shown that children who have a reading comprehension (but not a word recognition) problem are poor at making inferences from text: both inferences that connect up the ideas in the text (Oakhill, 1982; Oakhill, Yuill & Parkin, 1986) and inferences that incorporate knowledge about the world (Oakhill, 1983, 1984). The poor comprehenders' difficulty in making inferences might be partly due to an inadequate

Source: *British Journal of Developmental Psychology*, 1990, 8, 401–410.

understanding of the meaning of causal expressions (which signal that a causal link should be inferred).

There is some evidence that, although good and poor comprehenders do not differ in their ability to retrieve the meanings of individual content words, their semantic representations of word meanings may not be so detailed or subtle as those of skilled comprehenders (Oakhill, 1983; Perfetti, Hogaboam and Bell, cited by Perfetti & Lesgold, 1979). Since the semantics of causal connectives such as *because* involve a number of complex and subtle distinctions (Donaldson, 1986), it is reasonable to predict that comprehension of expressions using *because* will be a source of difficulty for poor comprehenders.

A weak grasp of the meaning of *because* might also contribute to another of the characteristics of poor comprehenders, namely their difficulty in integrating information from different clauses or sentences (Oakhill, 1982).

There is, as far as we know, no work on the understanding of causal expressions in skilled and less skilled reading comprehenders, though some unpublished data of our own provides evidence that skilled comprehenders are more likely to use *because* in their spoken productions. We asked groups of skilled and less skilled comprehenders, similar to the ones in the present experiment, to retell a story they had heard, which contained only the connectives *and* and *then*. We found that all children used *and* and *then* in their retellings, but only some used causal and temporal connectives, such as *because, so* and *when*. The skilled comprehenders introduced significantly more causal and temporal connectives into their retellings than did the less skilled group.

Studies which have explored children's understanding of causal expressions (without relating it to reading comprehension) have yielded conflicting results. Some studies have indicated that children do not understand *because* until at least the age of 7 years (e.g. Corrigan, 1975; Emerson, 1979; Piaget, 1926, 1928), whereas other studies have indicated that children have some understanding of the meaning of *because* by 5 years, or perhaps even as early as 3 years (e.g. Donaldson, 1986; Hood & Bloom, 1979; Trabasso, Stein & Johnson, 1981). This discrepancy between findings is probably largely attributable to methodological differences between the studies (see Donaldson, 1986, for a review). In general, those studies that found early acquisition of *because* used more naturalistic research methods than the other studies.

However, even those researchers who have found early understanding of *because* would concede that the level of understanding varies according to the type of causal sentence. In particular, children as old as 9 or 10 years sometimes have difficulty understanding *because* when it is used to introduce the evidence for a conclusion, whereas they have no difficulty understanding *because* when it is used to introduce the cause of an event (Donaldson, 1986; Trabasso *et al*., 1981). These findings suggest that any weaknesses in poor comprehenders' understanding of *because* may well be restricted to particular types of causal sentence.

In the present experiment, we explored skilled and less skilled comprehenders' understanding of two types of *because* sentence, using a task devised by Donaldson (1986). Donaldson made a distinction between explanations in the deductive mode and explanations in the empirical mode. The deductive mode occurs where a judgement or conclusion is justified in terms of some sort of evidence, for example:

We can tell that Mary is sad because she is crying.

In this sentence, *because* is used to introduce the evidence on which the conclusion is based. This type of sentence contrasts with sentences in the empirical mode:

Mary has a cold because she got soaked

where the *because* clause does not introduce evidence for a conclusion. The fact that Mary has a cold is, rather, the effect of her having got soaked. Thus, in the empirical mode, *because* introduces the cause of an event or state.

Donaldson showed her subjects pictures of simple causal sequences (e.g. Mary getting soaked; Mary sneezing) and then asked them questions (orally) to elicit either the empirical use of *because* ('Why does Mary have a cold?') or the deductive use ('How do you know Mary has a cold?'). She found that the youngest children in her study (5- and 8-year-olds) tended to interpret the deductive items as though they were empirical—they tended to answer *both* sorts of question with 'because she got soaked'. The 8-year-olds scored about 50 per cent correct on the deductive items, and even the 10-year-olds scored only 64 per cent. Subjects of all ages got most of the empirical items correct. In another condition, Donaldson used a completion task, which required the subjects to complete orally presented sentence fragments ('Mary has a cold because . . .' or 'We can tell that Mary has a cold because . . .'). The completions also showed an overwhelming tendency towards empirical interpretations in the younger children. They produced completions of the form 'she got soaked' for both types of item. Performance was poorer on the completion task than on the question task.

Although performance on the deductive items was poor overall, there were large individual differences: in each age group there were some children who gave consistently correct responses to deductive items. An interesting possibility is that these individual differences might correspond to individual differences in reading comprehension ability.

The present experiment aimed to explore the relation between reading comprehension ability and understanding of causal sentences in the empirical and deductive modes. We expected that the less skilled comprehenders would show a stronger tendency than the skilled comprehenders to interpret both the empirical and the deductive items as empirical.

Table 1 Characteristics of skilled and less-skilled comprehenders (standard deviations are shown in parentheses)

	Chronological age (years)	*Accuracy age (years)*	*Comprehension age (years)*	*Gates–MacGinitie (score/48)*
Less skilled	7.7 (0.32)	8.3 (0.81)	7.2 (0.30)	31.1 (8.5)
Skilled	7.8 (0.43)	8.3 (0.95)	8.7 (1.24)	31.9 (9.6)

Method

Subjects

Twenty-four subjects from two Brighton primary schools participated in the experiment. The children were divided into two groups, matched on tests of word recognition and reading vocabulary (Neale Accuracy and Gates-MacGinitie), but differing in performance on a reading comprehension test (Neale Comprehension). In order to select subjects, classes of 7- to 8-year-olds were screened initially using an adapted form of the Gates MacGinitie Primary B vocabulary test (Gates & MacGinitie, 1965). This test involves matching a series of pictures with a choice of four printed words per picture, and was administered to the children to provide a general indication of each child's sight recognition vocabulary. On the basis of this test, some children were selected and tested individually using the Neale Analysis of Reading Ability (form C) (Neale, 1966). This test provides age-related measures both of children's ability to read aloud words in context (accuracy age) and of their comprehension of short passages.

The 12 less-skilled comprehenders were chosen according to the following criteria: their reading accuracy age was average or above for their chronological age, but their comprehension age was at least half a year below their reading accuracy age. Twelve skilled comprehenders were chosen who were matched with the less skilled group for chronological age, Neale Accuracy age and Gates-MacGinitie scores, but who had relatively high comprehension scores for their ages. The characteristics of the two groups of subjects are shown in Table 1. The matching on Neale Accuracy scores was based on the regressed scores of the two groups, which takes into account the possibility that the two comprehension skill groups were selected from populations which differ in ability at word recognition. If this were the case, the groups might be found to differ in their reading-aloud scores if retested, owing to regression toward the mean scores of the populations from which they were derived (McNemar, 1962). More details of how the regressed scores were calculated can be found in Oakhill (1984). The groups did not

Table 2 Examples of materials

Empirical
1.
Pictures: Mary finds a mouse in her bed. Mary is hiding in the corner.
Questions: Mary is scared isn't she? . . . Why is Mary scared?
Completion: Mary is scared because . . .
2.
Pictures: John falls off his bike. John's leg is in plaster.
Questions: John has got a broken leg, hasn't he? . . . Why has John got a broken leg?
Completion: John's leg is broken because . . .
Deductive
1.
Pictures: Mary gets soaked. Mary is sneezing.
Questions: Mary has got a cold, hasn't she? . . . How do you know Mary has got a cold?
Completion: We can tell that Mary has got a cold because . . .
2.
Pictures: John bumps into Mary. There is a puddle of milk on the floor.
Questions: Mary spilt the milk, didn't she? . . . How do you know Mary spilt the milk?
Completion: We can tell that Mary spilt the milk because . . .

differ in Neale accuracy age or chronological age (both ts < 0.36). Their Neale comprehension scores did, however, differ significantly ($t(22) = 3.98$, (p) < .001). The skilled group was also selected so that their scores on the Gates-MacGinitie vocabulary test were similar to those of the less skilled group ($t(22) = 0.22$), an indication that their reading-aloud ability carried over to word–picture matching tasks and was not purely a decoding skill, which did not entail knowledge of word meanings.

Materials

The materials were derived from those used by Donaldson (1986, Experiment 6). Example materials are shown in Table 2. Each item was accompanied by two coloured pictures showing the events described (e.g. Mary out in the rain; Mary in bed, sneezing).

Design

There were 16 different materials, and each subject heard all 16, eight in the question task, and then a further eight in the completion task. In order to avoid confounding of materials with conditions, we derived four lists of materials such that, over the four lists, each material appeared once in each

mode (empirical/deductive) and once for each task (question/completion). Within each comprehension skill group, equal numbers of subjects were assigned to each list of materials.

Each list contained eight empirical and eight deductive items. Within the empirical and deductive items, half were presented for completion, and the other half were presented with questions. Deductive and empirical items occurred in a random order for each task (question/completion). In addition, half of the subjects in each group received the original instructions, as used by Donaldson, and half were given new instructions to induce them to attend more carefully to the information in the pictures (see Procedure section for more details). We included this condition because a pilot study with children of similar age and ability suggested that some children gave a very prompt answer to the questions, without listening carefully to the form of the question, and seemed to *expect* an empirical question.

Procedure

Each child was seen individually in a quiet room, and completed all 16 items in one experimental session. The items with questions were always presented before those for completion, and all items were presented orally. In the questions task, the experimenter first described the pictures for each item, for example: 'Mary gets soaked. Mary is sneezing', gave a tag question: 'Mary has got a cold, hasn't she?', and then asked a question: 'How do you know Mary has got a cold?' for the deductive items, or 'Why has Mary got a cold?' for the empirical items.

In the completion task, after a single non-causal practice item, the experimenter described the two pictures, as above, and then asked the subject to complete her sentence: 'Mary has got a cold because . . .' (empirical) or 'We can tell that Mary has got a cold because . . .' (deductive).

Half of the subjects in each group received the same intructions as Donaldson's, and half had instructions, in the deductive mode, that would lead them to attend more carefully to the information in the pictures. The latter group were asked, on these items: 'How do you know *from the picture* that Mary has got a cold? (question) and 'We can tell *from the picture* that Mary has got a cold because . . .' (completion).

Results

The children's responses were classified as either empirical or deductive: there were no 'don't know' responses. Separate analyses of variance were performed on the data for the two tasks: question answering and sentence completion. In both analyses there were two between-subjects factors: two types of instructions (original/new) × two levels of comprehension skill (good/poor), and one within-subjects factor: two modes (empirical or deductive).

Table 3 Mean numbers of correct question answers as a function of skill level, instructions and mode (empirical/deductive), max = 4

Mode	*Empirical*		*Deductive*		
Instructions	*Old*	*New*	*Old*	*New*	*Difference*
Less skilled	4.00	4.00	0.83	2.00	
	4.00		1.42		2.58*
Skilled	4.00	3.67	2.00	3.67	
	3.83		2.83		1.00*
Difference	0.17		1.41*		

* significant at the 5 per cent level (Tukey test).

Answers to questions

The results are shown in Table 3 as a function of mode, comprehension skill and type of instructions. There was a main effect of comprehension skill ($F(1,20) = 6.66$, (p) < .02). The skilled comprehenders were better overall. There was also a main effect of mode ($F(1,20) = 67.48$, (p) < .0001): empirical items were easier than deductive ones; and an interaction between comprehension skill and mode ($F(1,20) = 13.18$, (p) < .002). These effects including mode arose because, although the empirical mode was easier overall, the difference between the modes was much more marked for the less skilled than for the skilled comprehenders. Both groups did extremely well on the empirical items (in fact, the less skilled comprehenders got them all right). Although both groups performed more poorly on the deductive items, the less skilled comprehenders also performed significantly more poorly than the skilled comprehenders on these items (the difference was significant at the 5 per cent level on a Tukey test).

For clarity, the results as a function of instruction type and mode are shown separately in Table 4. There was a main effect of instructions

Table 4 Mean numbers of correct question answers shown as a function of instruction condition and mode (empirical/deductive), max = 4

	Mode		
Instructions	*Empirical*	*Deductive*	*Difference*
Old	4.00	1.42	2.58*
New	3.83	2.83	1.00*
Difference	0.17	1.41*	

* significant at the 5 per cent level (Tukey test).

Table 5 Mean numbers of correct completions as a function of skill level, instructions and mode (empirical/deductive), max = 4

Mode	*Empirical*		*Deductive*		
Instructions	*Old*	*New*	*Old*	*New*	*Difference*
Less skilled	4.00	4.00	0.33	2.50	
	4.00		1.42		2.58*
Skilled	3.83	3.83	1.67	3.50	
	3.83		2.58		1.25*
Difference	0.17		1.16*		

* significant at the 5 per cent level (Tukey).

($F(1,20) = 6.66$, (p) < .02), and this factor interacted with mode ($F(1,20) = 13.18$, (p) < .002). The instructions that drew the children's attention to the pictures produced better performance in the deductive mode (the difference was significant on a Tukey test at the 5 per cent level). Neither of the interactions involving comprehension skill was significant (both Fs < 1).

Sentence completions

The results are shown in Table 5 as a function of mode, comprehension skill and type of instructions. In general, these results are very similar to those for question answering. There was a marginal main effect of comprehension skill ($F(1,20) = 4.05$, (p) = .058): the skilled comprehenders were better overall. There was also a main effect of mode ($F(1,20) = 60.80$, (p) < .0001), and an interaction between comprehension skill and mode ($F(1,20) = 7.36$, (p) < .02). As in the question answering data, these effects including mode arise because, although the empirical mode was easier overall, the difference between the modes was much more marked for the less skilled than for the skilled comprehenders. Both groups did extremely well on the empirical items (as for question answers, the less skilled comprehenders got them all right) and less well on the deductive items. In addition, the less skilled comprehenders performed significantly more poorly than the skilled comprehenders on the deductive items (the difference was significant at the 5 per cent level on a Tukey test).

The results as a function of instructions and mode are shown separately in Table 6. As in the analysis of question answers, there was a main effect of type of instructions ($F(1,20) = 16.18$, (p) < .001), and this factor interacted with mode ($F(1,20) = 16.55$, (p) < .001). The instruction that drew the children's attention to the pictures produced better performance overall and, as

Table 6 Mean numbers of correct completions shown as a function of instruction condition and mode (empirical/deductive), max = 4

	Mode		
Instructions	*Empirical*	*Deductive*	*Difference*
Old	3.92	1.00	2.92*
New	3.92	3.00	0.92*
Difference	0.00	2.00*	

* significant at the 5 per cent level (Tukey test).

with questions, the improvement was confined to items in the deductive mode (the difference was significant on a Tukey test at the 5 per cent level). Neither of the interactions involving comprehension skill was significant (both Fs < 1).

Discussion

The main aim of the present study was to explore understanding of causal expressions using *because* in skilled and less skilled comprehenders. We compared their ability to understand usage of *because* in both the deductive mode, where a judgment or conclusion is based on some evidence, and in the empirical mode, where the link is between cause and effect. We found that, as predicted, the less skilled comprehenders performed in a similar manner to the younger children in Donaldson's (1986) study. They tended to interpret all items, including the deductive ones, as though they were empirical. The skilled comprehenders, by contrast, showed a more mature understanding of *because*, and were able to understand the deductive use much better, though not perfectly. These results indicate that poor comprehenders have difficulty handling *because* sentences, but that this difficulty is confined to the deductive mode.

The poor comprehenders' weak performance on deductive mode items is open to at least four interpretations, all of which have implications for reading comprehension, although further research is needed to investigate the possibilities outlined here.

One possibility is that the poor comprehenders' difficulties with deductive sentences are primarily linguistic. In particular, poor comprehenders may have an inadequate semantic representation of *because*, and hence they may not realize that *because* can be used to introduce evidence. Such a weakness in linguistic knowledge would have serious consequences for poor comprehenders' interpretation of deductive links in text. If they misinterpret *because*

in longer texts they will experience difficulty in keeping track of the causal structure of the text.

A second possibility is that the less skilled comprehenders are generally deficient in deductive reasoning skills. Indeed, Oakhill (1981: Experiment 13) showed that less skilled comprehenders are worse than skilled comprehenders at solving three- and four-term linear syllogisms. However, the overall pattern of results suggested that the less skilled comprehenders' problem lay, not in their deficient deductive reasoning skills *per se*, but in their ability to coordinate information in working memory to solve the more complex of the problems. This interpretation of their difficulties has certainly been borne out in more recent studies, which have shown that at least one source of the less skilled comprehenders' difficulty in integrating information and making inferences from text is their poor working memory skills (Oakhill, Yuill & Parkin, 1988). There are two reasons why the deductive items in the present experiment may have imposed heavier demands than the empirical ones on working memory. First, the deductive items were longer than the empirical ones. Second, as Donaldson (1986) points out, as well as asserting a deductive relation (e.g. 'We can tell that Mary spilt the milk because there is a puddle on the floor'), such sentences also presuppose an empirical relation (e.g. 'There is a puddle on the floor because Mary spilt the milk').

A third possible interpretation is that poor comprehenders' difficulties with deductive items reflect a lack of ability in relating evidence to conclusions to decide how a particular conclusion has been reached. In other words, their problems may be attributable to poor meta-cognitive skills. Knowing how a conclusion was derived is one of the meta-cognitive skills recently suggested as important in learning to read (Ryan & Ledger, 1984). Young children are often very poor at tasks that require them to monitor their own understanding (Markman, 1977, 1979), and similar deficiencies have been found in poor comprehenders (Baker, 1984; Garner, 1980; Oakhill *et al*., 1988). The relative paucity of such skills may make it difficult for poor comprehenders (and young children) to reflect on how they know that a particular conclusion holds. If children cannot tell 'how they know' a conclusion is true, then they will be unable to revise their conclusions and, therefore, will be unable to reinterpret text that they have misunderstood (see Oakhill *et al*., 1988).

A fourth possibility is that poor comprehenders are not lacking in linguistic or meta-cognitive capacities, but rather are failing to utilize these capacities appropriately. This interpretation has much in common with Bereiter & Scardamalia's (1982) argument that many of the difficulties that children have with written composition stem from 'executive problems', such as problems in directing attention to relevant aspects of the writing task. Bereiter & Scardamalia found that children's performance on writing tests could be improved by various forms of 'procedural facilitation', which they define as 'any reduction in the executive demands of a task that permits learners to make fuller use of the knowledge and skills they already have' (p. 52). Such

problems might also be related to working memory: a working memory deficit might result in executive problems, and the facilitation techniques might help because they reduce the load on working memory.

The modified instructions used for the deductive items in the present experiment could be regarded as a form of procedural facilitation, in that they directed the children's attention to the source of evidence (that is, the pictures). It is important to note that these new instructions did not actually give the children any specific hints about the *content* of the correct answer: the phrase 'from the picture' could have referred to either of the two pictures. The new instructions were merely cueing the children in to the *procedure* they should adopt to arrive at the correct answer. This minor change to the instructions resulted in a marked change in performance for both skill groups. Both in the case of question answering and completions, the new instructions considerably reduced the children's propensity to respond to the deductive items as though they were empirical. This finding suggests that at least part of the children's difficulty with deductive sentences is attributable to the executive problem of focusing attention on the relevants source of evidence. In particular, when the less skilled comprehenders are helped to overcome their predominant tendency to interpret *because* as though it were empirical, they do show some competence in handling deductive explanations.

In some of our other experiments, we have found that helping poor comprehenders to focus their attention on deductive processes can improve their reading comprehension. For example, training less skilled comprehenders in making links between words in a text, and what can be deduced from them, brings improvement in their performance on standardized, as well as specially designed, comprehension tasks (Yuill & Joscelyne, 1988; Yuill & Oakhill, 1988). These results again support the idea that the less skilled comprehenders' problem is not related to their basic deductive skills, but is related to a failure in meta-cognitive skills and their implementation. The present results suggest that future research might profitably explore whether training less skilled comprehenders in the identification of functions of different types of connective in text would help to improve their reading comprehension.

References

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology*, **38**, 289–311.

Bereiter, C. & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in Instructional Psychology*, vol. 2. Hillsdale, NJ: Erlbaum.

Corrigan, R. (1975). A scalogram analysis of the development of the use and comprehension of 'because' in children. *Child Development*, **46**, 195–201.

Donaldson, M. L. (1986). *Children's Explanations: A Psycholinguistic Study*. Cambridge: Cambridge University Press.

Emerson, H. F. (1979). Children's comprehension of 'because' in reversible and non-reversible sentences. *Journal of Child Language*, **6**, 279–300.

Garner, R. (1980). Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. *Journal of Reading Behavior*, **12**, 55–63.

Gates, A. I. & MacGinitie, W. H. (1965). Gates-MacGinitie Reading Tests. New York: Columbia University Teachers' College Press.

Hood, L. & Bloom, L. (1979). What, when and how about why: A longitudinal study of early expressions of causality. *Monographs of the Society for Research in Child Development*, vol. **44**, no. 6.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development*, **48**, 986–992.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, **50**, 643–655.

McNemar, Q. (1962). *Psychological Statistics*, 3rd ed. New York: Wiley.

Neale, M. D. (1966). *The Neale Analysis of Reading Ability*, 2nd ed. London: Macmillan Education.

Oakhill, J. V. (1981). Children's reading comprehension. Unpublished DPhil Thesis, University of Sussex.

Oakhill, J. V. (1982). Constructive processes in skilled and less skilled comprehenders' memory for sentences. *British Journal of Psychology*, **73**, 13–20.

Oakhill, J. V. (1983). Instantiation in skilled and less skilled comprehenders. *Quarterly Journal of Experimental Psychology*, **35A**, 441–450.

Oakhill, J. V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, **54**, 31–39.

Oakhill, J. V., Yuill, N. M. & Parkin, A. J. (1986). On the nature of the difference between skilled and less skilled comprehenders. *Journal of Research in Reading*, **9**, 80–91.

Oakhill, J. V., Yuill, N. M. & Parkin, A. J. (1988). Memory and inference in skilled and less skilled comprehenders. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds), *Practical Aspects of Memory*, vol. 2. Chichester: Wiley.

Perfetti, C. A. & Lesgold, A. M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. Weaver (Eds), *Theory and Practice of Early Reading*, vol. 1. Hillsdale, NJ: Erlbaum.

Piaget, J. (1926). *The Language and Thought of the Child*. London: Routledge & Kegan Paul.

Piaget, J. (1928). *Judgement and Reasoning in the Child*. London: Routledge & Kegan Paul.

Ryan, E. B. & Ledger, G. W. (1984). Learning to attend to sentence structure: Links between meta-linguistic development and reading. In J. Downing & R. Valtin (Eds), *Language Awareness and Learning to Read*. New York: Springer-Verlag.

Trabasso, T., Stein, N. L. & Johnson, L. R. (1981). Children's knowledge of events: A causal analysis of story structure. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation*, vol. 15. New York: Academic Press.

Yuill, N. M. & Oakhill, J. V. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, **2**, 33–45.

Yuill, N. M. & Joscelyne, T. (1988). Effect of organisational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, **80**, 152–158.

67

A QUASI-EXPERIMENTAL VALIDATION OF TRANSACTIONAL STRATEGIES INSTRUCTION WITH LOW-ACHIEVING SECOND-GRADE READERS

R. Brown, M. Pressley, P. Van Meter and T. Schuder

Second-grade, low-achieving students experienced a year of either transactional strategies instruction or highly regarded, more conventional second-grade reading instruction. By the end of the academic year, there was clear evidence of greater strategy awareness and strategy use, greater acquisition of information from material read in reading group, and superior performance on standardized reading tests by the transactional strategies instruction students. This is the clearest validation to date of educator-developed transactional strategies instruction.

Since Durkin's (1978–1979) seminal discovery that American students received little instruction about how to comprehend text, there have been extensive efforts to identify strategies that can be taught to students to increase their understanding of and memory for text. Early strategy research (for reviews, see Dole, Duffy, Roehler, & Pearson, 1991; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989) tended to focus on instruction of individual strategies and improvements in narrowly defined performances (e.g., improvement on standardized comprehension tests when reading strategies were taught). The typical research tactic taken in these studies was to teach one group of students to use a particular cognitive strategy while reading, often a strategy consistent with a theory of knowledge representation favored by the researcher, with control students left to their own devices to understand text as best they could. Through this approach, a relatively small number of individual strategies were proved effective in increasing

Source: *Journal of Educational Psychology*, 1996, 88(1), 18–37.

elementary students' comprehension of and memory for text (e.g., visualizing ideas in text, summarizing, and self-questioning). What the single-strategy investigations demonstrated was that if students were under exceptionally strong instructional control (i.e., they were told to carry out a particular strategy on a particular occasion), they could carry out strategies that would improve comprehension and learning. Seldom was generalized use of individual instructed strategies observed, nor was there evidence of generalized improvement in reading.

On the basis of what is now known about skilled reading, it is not surprising that improvement in reading has required more than instruction in single strategies. During the late 1970s and early 1980s, a number of analyses of skilled reading were conducted (e.g., Johnston & Afflerbach, 1985; Lytle, 1982; Olshavsky, 1976–1977; Olson, Mack, & Duffy, 1981; see Pressley & Afflerbach, 1995, for a summary). What became apparent was that skilled reading did not involve the use of a single potent strategy but rather orchestration of cognitive processes. This understanding—that skilled readers coordinate a number of strategies while reading—partially fueled researcher efforts to develop instructional interventions that involved teaching of multiple comprehension strategies (Baker & Brown, 1984).

A well-known researcher-designed, multiple-strategies instructional package was Palincsar and Brown's (e.g., 1984) reciprocal teaching. Palincsar and Brown taught students to apply four strategies to expository text as they read (generate predictions, ask questions, seek clarification, and summarize content). The students used these strategies in reading groups, with the adult teacher releasing control of the strategic processing as much as possible to the group. Palincsar and Brown's prediction, consistent with Vygotsky's (e.g., 1978) theory of socially mediated learning, was that participation in reading group discussions that involved predicting, questioning, seeking clarification, and summarizing would lead to internalization of these processes by group members. In fact, a month or two of such instruction produces noticeable improvement in the use of the focal strategies but only modest improvement on standardized reading tests (for a review, see Rosenshine & Meister, 1994).

In addition to the research of Palincsar and Brown (1984), there were other attempts to teach multiple comprehension strategies. Some involved presenting a large number of strategies quickly; these approaches typically failed to produce improvements in elementary-level readers' comprehension (e.g., Paris & Oka, 1986). Other interventions involved more intensive direct explanation and modeling of small repertoires of strategies; these approaches generally were more successful in improving reading (e.g., Bereiter & Bird, 1985; Collins, 1991; Duffy et al., 1987).

Many educators became aware of strategy researchers' instructional successes and began to import strategies instruction into classrooms. What became apparent, however, was that when strategies instruction was successfully

deployed in schools, it involved much more than the operations studied in the well-controlled experiments (Pressley, Goodchild, Fleet, Zajchowski, & Evans, 1989). This factor motivated Pressley and his colleagues to study extensively how elementary educators implemented comprehension strategies instruction in schools (see Pressley & El-Dinary, 1993).

After investigating several educator-developed programs, our research group proposed that effective elementary-level comprehension was "transactional" in three senses of the term (Pressley, El-Dinary, et al., 1992). First, readers are encouraged to construct meaning by using strategies that enable the linking of text content to prior knowledge, consistent with Rosenblatt's (1978) use of the term. Second, much of the strategies instruction occurs in reading groups, with group members using strategies to construct meaning together. As such, meaning-making is transactional in the sense that the constructed group understanding differs from the personalized interpretations individuals would have generated on their own, especially if they did not use strategies. This is consistent with the use of the term in organizational psychology (e.g., Hutchins, 1991). Third, the teacher's or group members' actions and reactions cannot be anticipated when the reading group uses strategies to construct interpretations. Rather, the responses of all members of the group (including the teacher) are determined in part by those of others in the group, which is a transactional situation according to social development researchers such as Bell (1968). Thus, group members co-determine each other's thinking about text. Because the strategy instruction the research group observed was so "transactional" in these three senses of the term, this type of instruction was called *transactional strategies instruction* (TSI; Pressley, El-Dinary, et al., 1992).

The short-term goal of TSI is the joint construction of reasonable interpretations by group members as they apply strategies to texts. The long-term goal is the internalization and consistently adaptive use of strategic processing whenever students encounter demanding text. Both goals are promoted by teaching reading group members to construct text meaning by emulating expert readers' use of comprehension strategies: to emulate how expert readers constructively respond when they need to understand challenging text (e.g., Pressley & Afflerbach, 1995; Wyatt et al., 1993). Expert readers are planful and goal-oriented when they read, combine their background knowledge with text cues to create meaning, apply a variety of strategies (e.g., from seeking the important information in text to noting details), monitor their comprehension, attempt to solve their comprehension problems, and evaluate their understanding and performance (e.g., Is the content believable? Is the piece well written? Am I achieving my goals?). The result is a personalized, interpretive understanding of text.

A variety of qualitative methods were used in the previous studies of transactional strategies instruction (see Pressley, El-Dinary, et al., 1992). These included (a) ethnographies; (b) interviews involving questions emanating

from Pressley, Goodchild, et al.'s (1989) tentative description of strategies instruction; (c) interviews constructed to illuminate observations made in program classrooms; (d) long-term case studies; and (e) analyses of classroom discourse. Although the TSI programs differed in their particulars, there were a number of common components (Pressley, El-Dinary, et al., 1992):

- Strategy instruction is long-term, with effective strategies instructors offering it in their classroom throughout the school year; the ideal is for high quality process instruction to occur across school years.
- Teachers explain and model effective comprehension strategies. Typically, a few, powerful strategies are emphasized.
- The teachers coach students to use strategies on an as-needed basis, providing hints to students about potential strategic choices they might make. There are many mini-lessons about when it is appropriate to use particular strategies.
- Both teachers and students model use of strategies for one another, thinking aloud as they read.
- Throughout instruction, the usefulness of strategies is emphasized, with students reminded frequently about the comprehension gains that accompany strategy use. Information about when and where various strategies can be applied is commonly discussed. Teachers consistently model flexible use of strategies; students explain to one another how they use strategies to process text.
- The strategies are used as a vehicle for coordinating dialogue about text. Thus, a great deal of discussion of text content occurs as teachers interact with students, reacting to students' use of strategies and prompting additional strategic processing (see especially Gaskins, Anderson, Pressley, Cunicelli, & Satlow, 1993). In particular, when students relate text to their prior knowledge, construct summaries of text meaning, visualize relations covered in a text, and predict what might transpire in a story, they engage in personal interpretation of text, with these personal interpretations varying from child to child and from reading group to reading group (Brown & Coy-Ogan, 1993).

Although the qualitative studies provided in-depth understanding of the nature of transactional strategies instruction programs, and a variety of informal data attested to the strengths of these programs (e.g., correlational, nonexperimental, and quasi-experimental comparisons conducted by school-district officials; see Brown & Pressley, 1994), what was lacking until the study reported here was conducted were formal comparisons on a variety of reading measures of students who received transactional strategies instruction versus more conventional instruction. This study begins to fill that gap. There were several important challenges to making such comparisons, however. One challenge was determining what should be measured. Reading

strategies instruction has tended to focus on gains on one or a few traditional measures of reading performance (Pressley, El-Dinary, et al., 1992). It became clear on the basis of the qualitative studies that transactional strategies instruction probably affected student cognition in a number of ways, however, with both short-term and long-term impacts (Pressley, Schuder, Teachers in the Students Achieving Independent Learning Program, Bergman, & El-Dinary, 1992).

A second challenge was that many indicators in the qualitative research conducted on transactional strategies instruction suggested that the effects of such an intervention appeared in the long term; that is, at a minimum, only after a semester to an academic year of such instruction (see Marks et al., 1993; Pressley, El-Dinary, et al., 1992; Pressley, Schuder, et al., 1992). A credible evaluation had to be long term. A constraint was that students often move in and out of schools at a high rate; thus, holding together large groups of students for several years was impractical. Our solution was to evaluate 1 year of transactional strategies instruction, because 1 year of intervention was all we believed could be completed in the participating district with an intact sample of students.

A third challenge resulted in our decision not to assign teachers randomly to conditions. Becoming an effective transactional strategies instruction teacher takes several years (e.g., El-Dinary & Schuder, 1993; Pressley, et al., 1991; Pressley, Schuder, et al., 1992). Thus, we felt we could not take any group of teachers and randomly assign them to transactional strategies instruction or control conditions. Moreover, we decided not to assign accomplished transactional strategies instruction teachers randomly to teach strategies versus some other approach. Because the teachers were committed to strategies instruction, we felt it was inappropriate to ask them to alter their teaching for an entire year. Our solution was to use a quasi-experimental design involving accomplished TSI teachers and other teachers in the same district, teachers with reputations as excellent reading educators whose instruction followed the guidelines of the district's regular literacy curriculum.

Before proceeding with a description of the formal methods in our study, we summarize some of the most important features of the Students Achieving Independent Learning (SAIL) program, the specific educator-developed transactional strategies instruction approach evaluated here. A description of SAIL will permit readers to understand our expectations in this quasi-experiment.

The SAIL comprehension strategies instructional program

The purpose of SAIL is the development of independent, self-regulated meaning-making from text. The program was developed over the course of a decade in one mid-Atlantic school district (see Schuder, 1993, for a history of SAIL and its evolution). SAIL students are taught to adjust their reading

to their specific purpose and to text characteristics (Is the material interesting? Does it relate to the reader's prior knowledge?). SAIL students are instructed to predict upcoming events, alter expectations as text unfolds, generate questions and interpretations while reading, visualize represented ideas, summarize periodically, and attend selectively to the most important information. Students are taught to think aloud (e.g., Meichenbaum, 1977) as they practice applying comprehension strategies during reading group instruction. For example, they reveal their thinking to others when they talk about their past experiences in relation to text. All of these reading processes are taught as strategies to students through direct explanations provided by teachers, teacher modeling, coaching, and scaffolded practice, both in reading groups and independently.

Direct explanations and modeling of strategic reasoning are critical components for preparing students to internalize and use strategies adaptively. These core components start the long-term process of helping students become more self-regulated and skillful readers. Direct explanations include providing students with information about the benefits of strategy use, as well as when and where to use strategies. In this excerpt, a SAIL teacher explained what is necessary to make good predictions:

T: We're going to set a couple of goals. So let's listen carefully and really really try to meet these goals by the end of the lesson. The first thing I want everybody to try to do is to make really good predictions. Can anybody tell me what a prediction is? S8?
S8: When you think what's gonna happen next.
T: When you try to guess what's going to happen next. If you're going to be a good predictor, how do you make good predictions? What do you need to have? S5?
S5: Enough information.
T: You have to have enough information in order for a good prediction to be made. Where can you find your information to get a prediction, where can you find it, S8?
S8: In the book?
T: In the book? You mean like, from what you've already read?
S8: Yeah.
T: S6?
S6: In your head.
T: In your head. Sometimes, a fancy word for that is background knowledge. In other words, knowledge means that you know. If you already know something about foxes, or about what a trot is, you might be able to even make a prediction about what the story is about. But maybe we should read a little bit into it to get a little more information to make sure we can make some good predictions.

Modeling, another critical SAIL component, does not consist solely of showing students how to use a strategy. Instead, SAIL teachers verbally explain their thinking and reasoning as they model appropriate use of strategies. In the following example, a SAIL teacher modeled her use of strategies, verbalizing her thinking as she applied a strategy in response to the demands of the text and her need to understand. However, before modeling, she explained to students why she was going to model: to help them observe how and why she used strategies to comprehend what she was reading:

> T: I'm gonna start this morning modeling like I usually do. . . . This is gonna be a real good opportunity for you to use a lot of your strategies. There are a lot of big vords in this story. Okay, so it's going to give us a chance to use some of our "fix-up kit" and it's gonna also give us a chance to use a lot of background knowledge, things that we already know from our own life, to help us understand what this story is about . . . ,

The teacher proceeded to read "Fox was a fine dancer. He could waltz, he could boogie, he could do the stomp." She then modeled her thinking for students:

> T: You know what? I'm thinking waltzing, boogieing, doing the stomp. I don't really know what the stomp is. But I'm thinking to myself that the stomp must be a dance because I do know that the waltz is a dance. That's when two people dance together. Because I used to see that on the Lawrence Welk Show. My grandmothers used to watch that a lot. And the boogie, well, I know that was a dance when I was in high school and that's when you move real fast. So, I'm thinking, the word *stomp*, I know that, well you can stomp your foot, and maybe that's what people do when they do the stomp. But I still think it's a dance. So that's what I'm gonna think, that. He could do the stomp, so that's a dance.

The teacher related information from her own experiences to text details. She used her prior knowledge to apply one of the "fix-up" strategies, guessing, when figuring out the meaning of an unknown word (i.e., *stomp*). Later in the same lesson, a student substituted a word when she came to a word she did not know. The teacher reminded students of the strategic reasoning she used when she initially verbalized her thinking. She then explained how readers can select different "fix-up" strategies (i.e., substituting a known word for an unknown one or relying on picture clues) to achieve the same purpose: to understand the gist of a passage.

> T: But, back over here, when I first said that there were some big vocabulary that had to do with the kinds of dances . . . we were talking about the fact that S1 substituted as a strategy and she was able to figure out . . . [what the word meant]. And then S6 over there gave her the real word, and then we found out that the real word wasn't that important because we could understand. Then I was thinking back to the fact that I didn't know what the stomp was and here I was looking at the picture clue. So, even though S1 was having trouble with the vocabulary, she could still get the gist of the type of dances by looking at the three pictures. There were three types of dances and there were three pictures.

In addition, SAIL students are taught multiple methods for dealing with difficult words, including skipping them, using context clues to determine the meaning of hard-to-decode and unfamiliar words, and rereading for more clues to meaning. Overreliance on any one strategy is discouraged. For example, skipping every unknown word can lead to comprehension failures, particularly if the skipped words are central to the meaning of the text. Instead, students are taught that skipping is just one of several strategies available to them when they encounter unknown words. When students are taught to ignore unknown words judiciously, skipping becomes a powerful problem-solving strategy for those who otherwise might linger too long over an undecodable word. In general, students are taught that getting the overall meaning of text is more important than understanding every word, so that difficult words sometimes can be skipped with little or no loss in meaning.

When SAIL instruction occurs in reading groups, it differs in a number of ways from more conventional reading group instruction: (a) Prereading discussion of vocabulary is eliminated in favor of discussion of vocabulary in the context of reading. (b) The almost universal classroom practice of asking comprehension-check questions as students read in group (e.g., Mehan, 1979) is rarely observed in transactional strategies instruction groups (Gaskins et al., 1993). Instead, a teacher gauges literal comprehension as students think aloud after reading a text segment. (c) There are extended interpretive discussions of text, with these discussions emphasizing student application of strategies to text.

Although reading group is an important SAIL component, the teaching of strategies extends across the school day, during whole-class instruction, and as teachers interact individually with their students. Reading instruction is also an across-the-curriculum activity. (For more detailed descriptions of SAIL, see Bergman & Schuder, 1992; Pressley, El-Dinary, et al., 1992; Schuder, 1993.)

One hypothesis evaluated here was that participating in SAIL would enhance reading comprehension as measured by a standardized test. A second was that there would be clear indications after a year of SAIL instruction

of students learning and using strategies. A third was that students would develop deeper, more personalized and interpretive understandings of text after a year of SAIL.

These hypotheses were evaluated with low-achieving second-grade students, a group targeted by SAIL: SAIL was designed originally for introduction to elementary students in either first or second grade who were at risk for reading failure. It is intended as a dramatically richer and more engaging form of instruction than the skill-and-drill approaches so often delivered to at-risk students (Allington, 1991). Thus, the evaluation reported here involved contrasting the achievement of low-achieving second-grade students who participated in SAIL with that of five matched groups of second-grade students receiving high quality, but more conventional, reading instruction.

Method

Participants

Teachers

The five transactional strategies instruction teachers served in the school district that developed the SAIL program; the five teachers in comparison classrooms were from the same school district. This district had garnered numerous national awards for excellence in instruction. Eight of the teachers taught second-grade classes. One SAIL teacher had a first/second-grade combination; one comparison teacher had a second/third-grade combination. All teachers were female. The SAIL teachers had 10.4 years of experience teaching on average; the comparison teachers averaged 23.4 years on average.[1] The five SAIL teachers exhausted the pool of second-grade teachers in the district with extensive experience teaching in the SAIL program (i.e., 3 or more years; range = 3–6 years). The comparison teachers were recommended by principals and district reading specialists, with nominations of effective second-grade teachers made on the basis of criteria such as the following: (a) They gave students grade-level-appropriate tasks; (b) they provided motivating learning activities; (c) they used classroom management well to avoid discipline problems; (d) they fostered active student involvement in reading; (e) they monitored student understanding and performance; and (f) they fostered academic self-esteem in students.

The comparison teachers were eclectic in their instructional practices, blending the whole-language tradition favored in the school district with elements of skill and other traditional forms of conventional reading instruction. For example, a teacher who stressed skills instruction sometimes integrated literature-based activities such as having students write in a response journal or read from a trade book (rather than a basal reader). A teacher who emphasized elements associated with a literature-based approach

also taught or reviewed phonics, word attack, and specific comprehension skills before or after reading. Some conventional instructors also taught a few strategies, like skipping unknown words, making predictions, or activating background knowledge. However, they did not teach a flexible repertoire of strategies using explicit, verbal explanations of thinking, elements characteristic of SAIL instruction. The comparison-group teachers had not participated in any SAIL professional-development activities.

All participating teachers were administered DeFord's (1985) Theoretical Orientation to Reading Profile (TORP), a 28-item instrument discriminating among teachers identifying with phonics, skills, and whole-language orientations. The scoring is such that those favoring phonics-based reading instruction score lower than those favoring skills instruction, who score lower than those identifying with whole language (scores range from 28 to 140). The SAIL teachers' mean score was 113 ($SD = 9.7$), and the comparison teachers averaged 73 ($SD = 7.2$), with the SAIL teachers differing significantly from the comparison teachers, dependent $t(4) = 6.24$, $p < .05$. (In the teacher comparisons, teachers were the units of analysis. Each teacher taught a reading group, with each group consisting of six students. SAIL and non-SAIL reading groups were matched on school demographic information. SAIL and non-SAIL students in the matched reading groups were paired on the basis of students' fall standardized test performances, described later in the Method section. As such, a correlated-samples analysis was conducted, because SAIL and comparison teachers were matched.) When the particular items of the TORP were examined, it was clear that the SAIL teachers had more of a whole-language orientation than the comparison teachers, who endorsed phonics and skills to a greater degree, smallest $|t|(4) = 4.88$, $p < .05$ for any of the three subscales. This finding was as expected, because SAIL encourages meaning-making as the goal of reading and discourages teaching of skills in isolation, consistent with whole language. Informal observations of the comparison teachers over the year confirmed that they were more eclectic in their approach to reading instruction than the SAIL teachers, incorporating a balance of whole-language, phonics-based, and skills-based instruction. Thus, their more balanced appraisal of the TORP items was consistent with our observations of their teaching.

At the beginning of the study, the 10 participating teachers were also administered a 25-item researcher-constructed questionnaire tapping their beliefs about teaching ($r = .94$; Cronbach's alpha calculated on participating teachers' responses). The questions involved responding to Likert-type statements (i.e., on a *strongly agree* to *strongly disagree* scale). For example, teachers who endorse transactional strategies instruction were expected to respond affirmatively to "The most important message to convey to students is that reading and thinking are inseparably linked," and "During instruction, teachers should ask story-related questions that have no precisely 'right' or 'wrong' answer." It was expected that SAIL teachers would disagree with

items such as "Worksheets that enable students to practice comprehension skills can be very useful for low-group students," and "During reading instruction, teachers need to guide students towards one best interpretation of a story." The responses were scored so that consistency with transactional strategies instruction would result in a low score (maximum score = 120; one item was discarded). The scores of the SAIL teachers ranged from 25 to 45 on this scale ($M = 36.8$, $SD = 9.5$), and comparison teachers' scores ranged from 62 to 76 ($M = 70.8$, $SD = 5.3$), a significant difference, dependent $t(4) = -8.84$, $p < .05$. In short, there were multiple indicators at the outset of the study that the SAIL teachers were committed to a different approach to teaching from the conventional teachers and that the SAIL teachers' beliefs about teaching were consistent with a transactional strategies instruction philosophy.

Students

Student participants were assigned to second grade but were reading below a second-grade level at the beginning of the year. They were identified as such through informal testing (teacher assessments involving reading of graded basal passages and word lists), results from assessments administered as part of the Chapter 1 program, and the previous year's grades and reports. Unfortunately, none of the assessments used by the school district to classify readers as weak at the beginning of the year were standardized measurements, although there was converging evidence from the informal measures that all participants experienced at least some difficulty reading beginning-level, second-grade material. Therefore, student mobility patterns, Chapter 1 status, ethnic and minority composition, size and location of schools, and overall performances on standardized tests were used to pair SAIL and comparison classes in the study. Moreover, because we did not have information about students' performance in previous years in any subject area and no formal test data existed for all these students, we administered a standardized achievement test. To attain greater comparability, a standardized achievement test was used to match students in each class as participants. A comprehension subtest of the Stanford Achievement Test (Primary 1, Form J; Grade level 1.5–2.5) was administered in late November or early December (depending on the class) of the school year. Administration of this test occurred at that point because only then did the teachers feel that participating students could function somewhat independently at the 1.5 grade level and thus not perform on the very floor of this measure. Unfortunately, this necessitated that the test be administered after SAIL teachers had introduced SAIL component strategies, so that it is not perfectly accurate to consider this a pretest.

Six students in each of the paired classes (i.e., a pair consisted of one SAIL class and one comparison class) were matched on the basis of their reading

comprehension scores. All of the children participating in the study spoke and comprehended English. In addition, the sample included no children experiencing severe attentional or behavioral problems. Only six students in one SAIL class met the eligibility requirements. Because students were matched on the basis of their standardized comprehension pretest scores, six matched pairs were selected for participation. Between the first and second semesters, 1 SAIL student and 2 comparison students in one pair of classrooms left their classrooms. Backup students were substituted, with no significant difference occurring between the newly constituted groups on the fall reading comprehension subtest.

There were five reading groups for the SAIL condition and five groups for the non-SAIL condition, each consisting of 6 students per group. Thus, in all comparisons between conditions, the reading group mean was the unit of analysis, with each unit consisting of the mean of the 6 designated students in each group.[2] With a maximum raw test score of 40 possible, the SAIL classes in the study averaged 22.20 on this measure ($SD = 6.85$) at the late fall testing, and the comparison classes averaged 22.67 ($SD = 5.89$), a nonsignificant difference (means per class analyzed), $t(4) = -0.59$, $p > .05$. Although not used for matching, the word skills subtest from the same standardized instrument was also administered (maximum score = 36 for the subtest), SAIL $M = 20.97$ ($SD = 2.76$), and comparison $M = 21.10$ ($SD = 3.40$), $t(4) = -0.10$, $p > .05$. The comparability of the paired groups is reflected well by considering their means and standard deviations on the fall Stanford Reading Comprehension subtest (see Table 1).

Although the 6 children from each classroom are referred to here as a reading group, their instruction varied through the year. First, reading was most often taught in homogeneous groups, although it also occurred during individualized and whole-class instruction. Second, participants often, but not always, remained members of the same homogeneous group over the course of the year (students who made great progress became members of another group). Because the SAIL program was offered to all children in the SAIL classrooms and the instruction in comparison classrooms did not resemble SAIL instruction, variable grouping did not pose a problem with respect to fidelity of treatment. The six participating children in each classroom did meet as a homogenous group for lessons that were formally analyzed, however. Even so, our use of the term *reading group* implies only that the 6 targeted children received either SAIL or conventional instruction daily, always within their classrooms, and frequently in small groups of students.

Design

This was an academic-year-long quasi-experimental study, carried out in 1991–1992. The reading achievement of five reading groups of low-achieving

Table 1 Stanford Achievement Test scores: matched SAIL and non-SAIL class means and standard deviations for word attack and reading comprehension

	Pretest					*Posttest*				
	SAIL		*Non-SAIL*			*SAIL*		*Non-SAIL*		
Matched classes	*M*	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*
T1 and T6										
Word Study Skills	19.33	6.35	16.33	3.08		28.00	4.05	22.33	5.75	
Reading Comprehension	20.83	6.94	20.33	7.92		36.83	4.36	26.83	9.28	
T2 and T7										
Word Study Skills	20.83	5.27	19.67	5.65		27.83	4.26	24.67	5.24	
Reading Comprehension	19.67	6.47	20.00	8.07		33.00	6.26	26.67	7.81	
T3 and T8										
Word Study Skills	18.67	4.72	23.33	2.73		23.67	2.80	22.83	6.01	
Reading Comprehension	15.67	4.63	16.83	6.24		30.67	1.63	26.00	9.94	
T4 and T9										
Word Study Skills	20.33	7.42	21.00	4.47		26.50	2.43	24.00	5.73	
Reading Comprehension	21.00	6.23	24.00	6.60		33.67	3.72	29.00	8.12	
T5 and T10										
Word Study Skills	25.67	2.50	25.17	5.04		29.50	5.79	26.17	5.60	
Reading Comprehension	33.83	7.28	32.17	6.88		36.83	2.40	35.17	4.22	
Group totals										
Word Study Skills	20.97	2.76	21.10	3.40	−0.10	27.10	2.19	24.00	1.53	3.98*
Reading Comprehension	22.20	6.85	22.67	5.89	−0.59	34.20	2.65	28.73	3.77	4.02*

Note: SAIL = Students Achieving Independent Learning. Maximum possible score on Word Study Skills subtest was 36, and on Reading Comprehension subtest, 40. SAIL and non-SAIL differences on Word Study Skills and Reading Comprehension pretests tested at $\alpha = .05$ (one-tailed).
* $p < .05$. one-tailed.

second-grade students receiving SAIL instruction was compared to the reading achievement of five groups of low-achieving second-grade students receiving instruction typical of second-grades in the district. Each of the 10 reading groups was housed in a different classroom, with each SAIL group matched with a comparison reading group that was close in reading achievement level at the beginning of the study and from a school demographically similar to the school representing the SAIL group. That is, there were five matched pairs of reading groups (6 low-achieving students per reading group), with one SAIL and one comparison reading group per pair.

The present study incorporated a quasi-experimental design in that we did not randomly assign teachers to conditions. Preparing teachers to become competent transactional strategies instructors is a long-term process; therefore, we felt we could not randomly assign teachers, provide professional development, and wait for teachers to become experienced in teaching SAIL in a realistic time frame. Also, the sample incorporated the largest cohort of experienced SAIL teachers in the school system. Therefore, we decided not to take SAIL teachers and randomly assign one group to teach SAIL and one group to teach another method. Even if we had access to a larger pool of SAIL teachers, we would not have asked them to alter for an entire year practices they were committed to.

The fact that SAIL teachers were committed to strategies instruction was not a concern; we felt that effective comparison teachers would be committed to the teaching practices they espoused as well. Although we might have attempted to identify potential comparison teachers in the buildings where SAIL teachers taught and randomly assigned students to teachers, we opted not to do this in favor of seeking the most competent second-grade comparison teachers that we could in the district. Because the comparison second-grade teachers did not serve in the same buildings as the SAIL teachers, random assignment of children to teachers was impossible. We believe the option we selected of matching reading groups taught by SAIL teachers with groups taught by teachers believed by the district administrators and reading consultants to be excellent second-grade reading teachers was a fair test of SAIL relative to highly regarded, more conventional reading instruction. We recognize that the use of a quasi-experimental design invites alternative explanations for results. However, we designed a study that was as close to experimental as possible by instituting as many precautions as we could.

Dependent Measures

The dependent measures are described in the order in which they were administered in the academic year. A summary of the measures appears in Table 2.

Table 2 Description of data sources for students

Data source	*When given*	*Why given*	*Description*
Strategies interview	October–November March–April	To assess SAIL and non-SAIL classes' awareness of comprehension and problem-solving strategies	Semistructured interview consisting of five base questions that were followed up with nondirective prompts; the questions were administered orally and individually to students.
Retelling questions	March–April	To assess students' retelling and sequencing of two stories presented by each teacher	Students individually were asked cued and picture-cued retelling questions.
Think-aloud task	May–June	To compare SAIL and non-SAIL classes' independent use of strategies during story reading; to determine if students were more text- or reader-based in their responses to probes	Students were stopped and asked "What are you thinking?" and other nondirective follow-up probes at four fixed points during story reading; students were questioned individually.
Standardized subtests of reading comprehension and word skills	November–December	To form comparable SAIL and non-SAIL reading groups by matching students using Stanford Achievement Test Reading Comprehension scores (fall administration)	Stanford Achievement Test, Primary 1, Form J
	May–June	To compare SAIL and non-SAIL classes on traditional, standardized, and validated measures of reading (fall and spring administration)	Stanford Achievement Test, Primary 1, Form K

Note: SAIL = Students Achieving Independent Learning.

Strategies interview

In October and November (i.e., when SAIL components were being introduced to SAIL students) and in March and April, a strategies interview was administered to all students participating in the study. This interview tapped students' reported awareness of strategies, as measured by the number and types of strategies they claimed to use during reading. We also hoped to assess whether students were aware of where, when, and why to use strategies.

Five open-ended questions (adapted from ones used by Duffy et al., 1987, for their study of strategies instruction with third-grade readers) were administered orally and individually to each participating student:

1. What do good readers do? What makes someone a good reader?
2. What things do you *do* before you start to read a story?
3. What do you *think* about before you read a new story?
4. What do you do when you come to a word you do not know?
5. What do you do when you read something that does not make sense?

These questions were presented in a different order for each student. If initial student responses were unclear or terse, the researcher probed for clarifications and elaborations.

Story lessons and retelling questions

In March or April (depending on class schedule), two stories were read by all participating reading groups. The instruction and interactions that occurred during reading were recorded as these stories were read, and they were analyzed to document differences in instruction in the SAIL and non-SAIL reading groups. (See the Appendix for a description of two SAIL and two non-SAIL lessons serving as a general comparison of SAIL and conventional group instruction.) A descriptive analysis of the lessons revealed that SAIL teachers more often gave explicit explanations, verbalized their thinking, and elaborated explicitly and responsively in reaction to students' comments and actions. Non-SAIL teachers more frequently than SAIL teachers provided information or instruction to students without stating a purpose, gave answers to students when they had difficulty reading or answering questions, drilled students on their learning, and praised and evaluated their performance. Both groups activated students' background knowledge, reviewed previously learned information, and guided students through their difficulties to about the same extent (Brown, 1995a, 1995b).

After the lesson was conducted, each student was asked to retell the story to a researcher, followed by a task requiring students to sequence pictures corresponding to events in the story. The primary purpose of this measure was to assess students' recall of text details, although we thought students might include interpretations in their retellings of story content as well.

All reading groups in the study read the same two illustrated stories. "Fox Trot" was a chapter in a popular children's trade book, *Fox in Love* (Marshall, 1982); "Mushroom in the Rain" (Ginsburg, 1991) was from the Heath Reading Series, Book Level 1. The readability for the 341-word "Fox Trot" was 2.4; the readability for the 512-word "Mushroom in the Rain" was 2.2 (Harris–Jacobson Wide Range Readability Formula; Harris & Sipay, 1985, pp. 656–673).

In "Fox Trot," the main character, Fox, decides to enter a dance contest. He asks two friends to be his dance partner, but they refuse. They suggest that Fox ask Raisin, but Fox is reluctant to do so because Raisin is angry with him. Nevertheless, he asks and she agrees. They practice hard and dance quite well together. On the day of the contest, Raisin gets the mumps. Fox returns home and despondently sits in front of a blank TV. Then he decides to teach his sister the dance steps. They rush to the contest and claim second prize.

In "Mushroom in the Rain," an ant seeks shelter from a storm. She squeezes herself into a small mushroom. A butterfly comes by and asks if he can escape the rain as well, with the ant allowing the butterfly to crowd in. Then comes a mouse and a bird, with the crowding in the mushroom increasing. A rabbit then arrives, who is being chased by a fox. The others hide the rabbit in the mushroom. Once the fox leaves and the rain stops, the ant asks the others how they managed to fit under the mushroom. A frog, sitting on top of the mushroom asks, "Don't you know what happens to a mushroom in the rain?" In the version of the story used in the study, the answer was not provided to the children but was left for them to infer.

These stories were selected because they provided ample opportunity for diverse interpretations and personal responses. They were on the school system's approved list and approved by the participating teachers as appropriate for a single lesson for weaker second-grade students in the spring.

All decisions about how to present the stories were made by the teachers. However, they were asked to present each of these stories in one morning lesson that was not to exceed 55 min in length. They were consistent with this request, with the mean SAIL lesson lasting 43.40 min (SD = 7.83) and the mean comparison-group lesson lasting 35.50 min (SD = 13.34). (Three of the matched pairs of reading groups read "Mushroom in the Rain" first; two pairs read "Fox Trot" first.) Generally, SAIL lessons are lengthier because negotiating interpretations, explaining and modeling strategies, thinking aloud, and selecting and using "fix-up" strategies while reading are time-consuming activities, particularly when they are compared with some activities in conventional reading lessons (e.g., answering skill-and-drill and literal comprehension questions). The lessons were videotaped to allow a manipulation check to ascertain that teaching in the SAIL groups was different as expected from teaching in the comparison reading groups (described in the Results section).

Approximately 2 hr after each lesson was over, each of the 6 students in the reading group was interviewed individually. First, students were asked to retell the story:

> Pretend that you are asked to tell the story to other kids in the class who have never heard the story before. What would you tell them happened in that story? Can you remember anything else? (Adapted from Golden, 1988)

This interview was followed by a cued, picture retelling task. Students were asked to sequence six scrambled pictures taken directly from the story. The students were then informed that sometimes pictures assist in aiding recall of stories, and they were asked to use the pictures to prompt recollection of story content.

Think-aloud measure

In May or June, students read a 129-word illustrated Aesop's fable, "The Dog and His Reflection," selected from a trade book (Miller, 1976). The readability for this story was 3.9 (Harris & Sipay, 1985), making it challenging for the students.

In the story, a dog steals a piece of meat from the dinner table. He runs into the woods and starts to cross a bridge. When he chances to look down, he sees his reflection in the water. Thinking his reflection is another dog with a larger cut of meat, he decides to seize the dog's chop. When he opens his mouth, his own piece of meat plunges into the water. Consequently, the dog ends up with nothing at all.

The students met with the researcher individually for this task. Students were stopped at four points in the reading of the Aesop fable and asked to report their thinking. If a student had difficulty reading a segment, the first question posed was, "What do you think happened on this page?" Otherwise, the student was asked first, "What are you thinking?" Both questions primarily focused on content, with the "What are you thinking?" probe designed to be open-ended enough to elicit interpretive remarks and opinions about the fable, although we expected students to recount story details as well. Thus, the first purpose of the think-aloud task was to supplement the story-recall task. Unlike the recall questions that were designed primarily to assess memory for story details, the more open-ended, think-aloud prompt was used to examine students' understandings and interpretations of text.

The other purpose of this measure was to supplement the strategies interview. Although the strategies interview revealed whether students talked about strategies, it did not indicate whether students used them on their own when reading. One limitation of the strategy interview was that students might memorize information repeated by their teachers without being able to

translate that knowledge into practice. Therefore, a task was designed to observe whether students actually used comprehension strategies when reading. Our intent was not to have these young students report directly on their thinking processes while reading. Instead, we observed whether students would use comprehension strategies when they were not cued to do so.

When students offered unelaborated responses to initial questions, open-ended follow-ups were asked (see Garner, 1988, p. 70), such as, "Can you tell me more?" or "Why do you say that?" Sometimes an unelaborated comment was echoed back to the student in the form of a question. Thus, after a student remarked that a dog stole a piece of meat from his master's table, the researcher asked, "What do you think about the fact that the dog stole a piece of meat from his master's table?" For every text segment, before the student moved on to reading the next segment, the researcher asked, "Is there anything you could say or do before reading on?"

Stanford Achievement Test subtests

In May or June, students took the Stanford Achievement Test (The Psychological Corporation, 1990), Form K, Reading Comprehension and Word Study Skills subtests. Standardized tests traditionally have been used as measures of reading performance in strategy experiments. Therefore, in addition to the other measures, students were compared on a conventional measure of reading achievement.

The Reading Comprehension subtest consists of two-sentence stories, comprehension questions on short passages, and sentence-completion items that form short stories. The Word Skills subtest includes questions pertaining to structural analysis (e.g., compound words, inflectional endings, contractions) and phonetic analysis (e.g., consonants and vowels). The comprehension test was administered first to all students, followed by the word skills test. The alternate-forms reliability for the full scale scores of Forms J (administered in the fall) and K was .89.

Results

Every hypothesis tested here was one-tailed, and each was an evaluation of whether SAIL instruction produced better performance than the comparison instruction. Most means appeared in only one hypothesis test, and hence, $\alpha < .05$ was the Type 1 error probability selected for all hypotheses (Kirk, 1982, for this and all references to statistics). For the standardized test data and strategies interview data, the simple effect of condition within time of testing was evaluated in the fall, as it was in the spring. The Time of Testing $\times$ Condition interaction was also tested. The hypothesis-testing approach taken here was conservative, providing high power for detection of large effects only (Cohen, 1988). For each dependent variable, the same overall

Type 1 error probability would have occurred if we had analyzed the data within a 2 × 2 analysis of variance structure.

All tests were based on the reading group mean as the unit of analysis (i.e., $n = 5$ groups for the SAIL condition, and $n = 5$ groups for the comparison condition, each consisting of 6 students per group), because individual scores within reading groups were not independent (see Footnote 2). Finally, all *t* tests were dependent *t* tests that were based on the 5 matched pairs, with one SAIL and one comparison group to a pair, with pairings determined by demographic information and by the reading groups' fall standardized comprehension performances, as described earlier.

For every dependent *t* test involving student posttest performance, an exact permutation test was also conducted. In all cases except three, performance in SAIL classes significantly exceeded performance in the comparison classrooms, $p = .03125$ (one-tailed) for the permutation test. In the two exceptions reported in the main text (i.e., the pretest-to-posttest gain on the standardized comprehension measure and the pretest-to-posttest gain on the strategies interview: word attack strategies), the gains for one of the 5 SAIL and non-SAIL pairs were identical. The SAIL classes exceeded the non-SAIL classes in the other pairs for both measures ($.03 < p < .07$, one-tailed). The third exception was in a supplementary analysis.[3]

In general, the results are reported in the order in which dependent measures were described in the Method section, which parallels the order of data collection in the study.

Fall–Spring strategies interview

The interviews were designed to determine whether SAIL and comparison students would differ in their awareness of strategies, operationalized as the number of strategies they claimed to use during reading. Two raters scored 20% of the interviews, with an overall 87% agreement for the strategies named by students. Only one of the two raters scored the remainder of the interviews.

A strategy was scored as mentioned if it was named in response to any of the interview questions. Any strategies mentioned by students were recorded, even if they were not strategies taught in the SAIL program. The comprehension strategies mentioned included the following:

Predicting: Guessing what will happen next
Verifying: Confirming that a prediction was accurate
Visualizing: Constructing a mental picture of the information contained in the text segment
Relating prior knowledge or personal experiences to text: Making an association between information in the text and information in the readers' head

Summarizing or retelling: Saying the most important information (summarizing) or restating in one's own words everything that occurred in the text segment just read
Thinking aloud: Verbalizing thoughts and feeling about text segments just read
Monitoring: Explicitly verbalizing when something just read does not make sense
Setting a goal: Deciding a purpose prior to reading, including decisions about both expository and narrative texts
Browsing or previewing: Flipping through the story, glancing at the pictures, or reading the back cover to get ideas about the story
Skipping: Ignoring a problematic part of text and reading on
Substituting or guessing: Replacing a difficult part of text with something else that appears to make sense and maintains the coherence of the text segment
Rereading: Returning to a problematic segment of text
Looking back: Looking back in the text for information that might help in understanding a difficult-to-understand part of text
Clarifying confusions: Asking a specific question to resolve a comprehension problem
Asking someone for help: Asking another student or the teacher for help with the confusing section of text

The following strategies for attacking unknown or difficult words were mentioned:

Skipping: Ignoring a problematic word and reading on
Substituting or guessing: Replacing an unknown word with another word that appears to make sense or that maintains the coherence of the text segment
Rereading: Returning to a problematic word
Looking back: Looking back in the text for information that might help in understanding a difficult-to-understand word
Using picture clues: Looking at pictures in the story to help determine the meaning of an unknown word or difficult-to-understand piece of text
Using word clues: Relying on the surrounding text to help decide the meaning of an unknown word or difficult-to-understand piece of text
Breaking a word into parts: Seeing if there are recognizable root words, prefixes, or suffixes contained within the larger word
Sounding out a word: Applying knowledge of phonics to the decoding of the word
Asking someone for help: Asking another student or the teacher for help with the confusing word

The comprehension and word-level strategies reports are summarized in Table 3. The means reported in the table are reading group means (i.e., a mean frequency of strategies reported for each reading group in the study was calculated on the basis of individual reading group members' reports, with each of the Table 3 means and standard deviations calculated on the basis of five reading group means). With respect to reports of comprehension strategies, there was no significant advantage for the SAIL students in the fall, shortly after the program had begun. By spring, however, as expected, the SAIL groups reported many more strategies than the comparison groups. In the spring, only SAIL students reported visualizing, looking back, verifying predictions, thinking aloud, summarizing, setting a goal, or browsing. Although during the spring interview, comparison-group students mentioned predicting, using text or picture clues to clarify confusions, making connections between text and their background knowledge and experiences, asking someone for help, skipping over confusing parts, and rereading, the mean frequency of such reports was always descriptively lower for them compared to the SAIL students. The SAIL and comparison groups mentioned monitoring and guessing approximately equally on the spring interview.

There were qualitative differences in students' responses to the strategy interview questions as well. When asked, "What do good readers do?" SAIL students responded more frequently than non-SAIL students that good readers use comprehension strategies, apply problem-solving strategies, and think. Both groups mentioned that good readers read abundantly, practice frequently, read well, and read for enjoyment. In response to questions about what students do or think before they read a story, students in both groups said they made predictions. However, SAIL students tended to predict what would happen in the story, whereas non-SAIL students predicted whether the story would be too difficult or whether they would like it. When asked, "What do you do when you read something that does not make sense?" students in both groups frequently mentioned they would skip or reread a confusing section; however, SAIL students cited these strategies more frequently.

With respect to word-level strategies, the SAIL students reported more strategies than the comparison-group participants, even during the fall interview (see Table 3). In the fall, SAIL students mentioned skipping words (see Footnote 3), substituting or guessing, using picture or word clues, rereading, and breaking words into parts descriptively more often than did comparison students. There was slightly more mention of sounding out of words in the comparison condition in the fall. The introduction to SAIL from the very start of school probably accounts for this fall difference in word-level strategies reports. By the spring, all of the word-level strategies were being mentioned by SAIL students. In contrast, the only word-level strategies mentioned consistently by more than 1 student per comparison reading group were skipping an unknown word, sounding a word out, rereading, and asking someone for help.

Table 3 Means and standard deviations for number of comprehension and word-level strategies mentioned in the fall and spring strategies interviews

	Fall					*Spring*				
	SAIL		*Comparison group*			*SAIL*		*Comparison group*		
Strategy	*M*	*SD*	*M*	*SD*	*t(4)*	*M*	*SD*	*M*	*SD*	*t(4)*
Comprehension	0.79	0.45	0.88	0.44	0.58	4.20	0.86	1.25	0.48	9.53
Word level	2.16	0.79	1.15	0.28	3.52	3.22	0.63	1.68	0.37	4.83

Note: SAIL = Students Achieving Independent Learning. With the exception of comprehension data in fall interviews, SAIL data were significantly greater than comparison data, $p < .05$, one-tailed.

We also tested whether SAIL students made greater gains in self-reported awareness of strategies over the course of the year. The one-tailed interaction hypothesis test (e.g., fall-to-spring increase in students' strategies scores by condition) was significant, as expected, for both the comprehension strategies, $t(4) = 7.72$, and the word-level strategies, $t(4) = 2.64$.

In general, SAIL students provided more elaborate responses to post-measure questions. For example, this rich spring interview was provided by a student in a SAIL class:

R: What do good readers do?
S: [Good readers have] lots of expression. They do think-alouds.
R: They do think-alouds. Okay. What do you mean by that?
S: Well, they tell people what they think is going on in their own words in the story.
R: Uh huh. What other things do good readers do?
S: Well, I'm an expert reader. And what I do is I skip. But, well, skipping isn't always great because sometimes you need to get the gist of the story. Cause if you always skip, you can't get the meaning of the story.
R: So you can't be skipping everything in the story. . . .
S: [I also do] substituting, and sounding things out is a very good strategy [*sic*] . . . and, um, looking back is a good strategy.
R: Looking back. . . . Why is looking back a good strategy?
S: Because like if I got stuck on a word, like, uh, it might be back on the story. . . . But sometimes it isn't.
R: Are there any other things good readers do? Are there any other strategies good readers use?
S: Guessing too. Picture clues are very good. . . . "The Cat and the Canary" has beautiful illustrations, and we think it should have a Caldecott medal 'cause of the picture clues. I looked there and the word "suddenly" came up because the picture clues just looked like "suddenly."
R: What things do you *do* before you start to read a story?
S: I look at the title, and I look first at the pictures.
R: Why do you do that?
S: Because that can give me information about what the story is about. But when I make predictions it's not always right. We don't get upset because it's not right. We just know that it's not right and then something goes off in our mind telling us that we should make another one.
R: What do you *think* about before you read a new story?
S: I think about whether it might be good or bad. . . .
R: How would you tell if it were good or bad?

S: If I were alone at home, I would look at the first pictures and start reading the first page and then I get ideas.
R: Okay, then, what might you do after you read the first page and get ideas?
S: I have a think-aloud in my mind that would tell me what the story might be about.
R: What do you do when you come to a word you do not know?
S: I use picture clues, I guess, look back, and sometimes I reread the sentence.
R: What do you do when you read something that does not make sense?
S: I read the sentence very slowly to see if I skipped a word.
R: Hmmm, what else do you do?
S: Sometimes I just skip it and go to the next line.

The following interview is representative of the type of responses given by non-SAIL students to the interview questions. Although some of the same components are apparent (particularly with word-level strategies), the student's responses are less elaborated:

R: What do good readers do?
S: They read a lot of books.
R: Anything else?
S: Nope.
R: What things do you *do* before you start to read a story?
S: Read the title.
R: Read the title. Why do you read the title?
S: Because, when, . . . if you don't read the title you won't know what it's about.
R: What do you *think* about before you read a story?
S: It might be tales.
R: It might be tales . . . what do you mean? Tell me a little more. . . .
S: Like, it might be funny.
R: Ah, so it might be funny . . . and how might you find that out? You haven't started reading it yet.
S: You might ask someone who read the story.
R: And what do you do when you come to a word you do not know?
S: You could ask your mother.
R: Is there anything else you could do?
S: I skip and then read the other words and then when you have finished the sentence, you could go back to that letter and you can sound it what it is.

R: What do you do when you read something that does not make sense?
S: You might read the word that you don't know and you're not sure what it is.
R: Anything else?
S: No.

Although the SAIL students mentioned a descriptively greater number of comprehension and problem-solving strategies than non-SAIL students, their responses did not reflect a high degree of complex reasoning about why using strategies is so beneficial. Students exhibited some rudimentary knowledge of when to use strategies appropriately: they were able to respond to questions about what they did or thought before reading and when encountering problems. Also, students were starting to understand that strategies could be used flexibly, especially for problem solving. Mentioning several strategies may have suggested some prerequisite understanding about the adaptive use of strategies. However, students' responses typically did not indicate precise conditions under which certain strategies could be applied effectively.

In summary, by spring the SAIL students definitely reported more comprehension and word-level strategies during the interview than did comparison-group students. That SAIL students were already reporting more word-level strategies in the fall than comparison students probably reflected the effects of the first month or two of instruction in the program. By spring, every strategy except two was mentioned descriptively more often in the SAIL than in the comparison group. The exceptions were sounding it out (which was consistent with the teaching philosophy of the comparison teachers) and asking for help with a word (which is difficult to construe as a strategy associated with independence in reading). Most important, SAIL students learned more about comprehension and word-level strategies over the year than comparison students. However, in general, this information concentrated more on awareness and naming of strategies than on deep understanding of how strategic reasoning works. The results suggest that fully self-regulated thinking is the product of years of development. Perhaps, too, the questions were neither precise nor concrete enough to probe the understanding of young children in an in-depth manner. Furthermore, the students may not have been able to verbalize knowledge of their own strategic processing (Pressley & Afflerbach, 1995).

Spring story lessons

Teaching of the lessons

The March–April lessons were transcribed from the videotape records, with the transcriptions read by four raters who were "blind" to condition.[4] One

rater was a SAIL program developer, and the other three were graduate students familiar with transactional strategies instruction and the SAIL program in particular. The program developer correctly classified 9 of the 10 SAIL lessons as consistent with the intent and original vision of the SAIL program; this rater definitely was sensitive to whether teachers explained and modeled strategic processes and encouraged interpretive construction of text meaning by students through use of comprehension strategies. The curriculum developer looked for evidence that the teachers thought aloud in their lessons and coached students to engage text actively (i.e., to relate text content to prior knowledge as well as to apply other strategies as appropriate). He classified all of the comparison lessons as not consistent with the SAIL approach and, in fact, not even close to being consistent with SAIL. The three graduate students correctly classified lessons as SAIL or non-SAIL for 59 of the 60 ratings made. Thus, there were clear instructional differences between the SAIL and non-SAIL classrooms during the March–April lessons.

Two raters reviewed the lessons (one rater was "blind" to condition) for evidence of strategies teaching, with interrater agreement of 85% and disagreements resolved by discussion (see Footnote 4). Collapsing across the two lessons observed for each teacher, a mean of 9.20 (SD = 1.92) different comprehension strategies were observed in the SAIL lessons compared to a mean of 2.00 (SD = 0.71) in the comparison lessons, $t(4) = 7.43$. Predicting, relating text to background knowledge, summarizing, and thinking aloud were observed in all SAIL groups. Only relating to background knowledge was observed in all comparison groups. In no SAIL group were fewer than seven of the comprehension strategies taught; in no comparison group were more than three observed.

On average, again collapsing across each participating reading groups' two lessons, 4.80 (SD = 0.45) word-level strategies were observed in the SAIL groups, and 4.00 (SD = 0.71) were documented in the comparison reading groups, $t(4) = 4.00$. Using semantic context clues and using picture clues were observed in all SAIL groups; using picture clues and sounding words out were observed in all comparison classrooms. The range of word-level strategies was between 4 and 5 in the SAIL groups and between 3 and 5 in the comparison groups. Thus, one important indicator that the instruction in the SAIL groups differed from comparison instruction was that there was more strategies instruction in the SAIL groups. The difference was much more striking with respect to comprehension strategies, however.

Student recall of stories covered in lessons

The recall protocols were analyzed using a modified analytic induction approach (Goetz & LeCompte, 1984); that is, coding categories emerged from analysis of the data. However, identification of categories also was highly

informed by the work of O'Flahavan (1989) and Eeds and Wells (1989). In this study, only the results of the literal and interpretive analyses are presented, because only they relate directly to the stated hypotheses.

The full categorization scheme and analysis can be found in the work of Brown (1995a, 1995b). Both "Fox Trot" and "Mushroom in the Rain" were parsed into idea units, a variant of the T unit (Hunt, 1965). Loosely defined, an idea unit is a segment of written or oral discourse that conveys meaning, consisting of a verb form with any associated subject, object, and modifiers. Length or grammatical structure does not determine whether a segment is coded as an idea unit; what counts is whether the unit is meaningful. Interrater agreement was calculated for 20% of the recalled stories (see Footnote 4). It was 89% for classification of the protocols into idea units of various types (e.g., literal, interpretive).

A first issue addressed was whether SAIL students recalled more interpretive idea units than comparison students. These remarks reflected students' relating of background knowledge to text. Interpretive ideas were not explicitly stated in the text or in the pictures but did not contradict information in the text or pictures. For instance, for the Mushroom story, "He wanted to be dry" was scored as an interpretive remark. (The text had said, "One day an ant was caught in the rain. 'Where can I hide?' he wondered. He saw a little mushroom peeking out of the ground in a clearing and he hid under it.") Also, the comment, "But they tricked him," was scored as an interpretive unit for the Mushroom story. (The corresponding text was, "How could a rabbit get in here? Don't you see there isn't any room," said the ant. The fox turned up his nose. He flicked his tail and ran off.") As a third example, one not corresponding to any specific part of the Mushroom story, the remark, "And it was the only place to keep him dry," was coded as an interpretive remark because it was a conclusion that did not contradict anything in the text.

For the Mushroom story, SAIL groups averaged 6.12 interpretive units per student ($SD = 1.54$), which exceeded the corresponding figure of 4.48 in the comparison groups ($SD = 1.70$), $t(4) = 2.99$. For "Fox Trot," SAIL groups averaged 5.58 interpretive units per student ($SD = 1.63$), which exceeded the corresponding figure of 3.84 in the comparison groups ($SD = 1.63$), $t(4) = 2.97$.

In the example below, a SAIL student interjected a personalized interpretation into his retelling of story events. Interpreting occurred even though the task was not designed to elicit such information. The text stated that the frog asked the other animals if they knew what happened to a mushroom when it rained. He then hopped away, laughing. The student recall included the following response:

> S: In the story, um um, the frog was just laughing because it was a miracle that came true. And the frog was laughing, the frog was

> laughing at them. And then really really when he was talking he said, "Don't you know what happens when it rains over a mushroom? And they they didn't know. They thought it was just a miracle, and when it was getting bigger it looked like a sleeping cap. So I think it was going wider and wider, and afterward when the sun came out and the fox was like an evil spirit, it went away. Um, they came, they came right out, and the mushroom was so big they didn't know what happened.

After the retelling was over, the researcher, curious about the origins of the student's interpretation, asked why he thought the fox was an evil spirit. The student replied, "Because it's like you know, the movies. And once there's this evil spirit and it's dark and nothing happens right. And once you kill it, the evil spirit, or if it goes away, and then it turns back into a good life." Thus, the student used his personal knowledge accrued from viewing movies to generate a unique interpretation that entered into his retelling.

In addition to scoring interpretive recall, we evaluated literal recall of ideas represented either in the stories or in the accompanying pictures. For example, one idea unit represented explicitly in the Mushroom story was, "He hid under it." If the student recalled this idea unit or a paraphrase of it, the student was scored as having recalled the unit. In "Fox Trot," there was a picture of Carmen and Dexter looking through a window, watching Fox dance. One idea unit was scored as recalled if the student reported something like, "His friends were looking at him dance from the window."

For the Mushroom story, SAIL reading groups recalled an average of 17.64 (out of a maximum of 79) literal idea units per student ($SD = 3.95$), which did not exceed literal recall in the comparison groups, who averaged 15.82 units ($SD = 1.31$), $t(4) = 1.10$. For "Fox Trot," however, SAIL recall ($M = 12.26$ out of a maximum of 59 units; $SD = 2.72$) exceeded comparison-group recall ($M = 8.38$, $SD = 2.94$), $t(4) = 2.60$.

In summary, SAIL students were significantly more interpretive in their recalls than comparison students, consistent with our expectations. Even though the questions called for literal recall of story content, SAIL students were more interpretive. This result is consistent with the conclusion that an interpretive propensity is internalized by TSI students. There were not strong expectations about the literal recall of the stories on the basis of condition, for we recognized that the comparison teachers covered the literal content of stories very well in their lessons. Even so, the students in the SAIL groups recalled more literal information than students in the comparison groups, although the difference favoring the SAIL students was significant for only one story.

One explanation of the story-recall results is that the SAIL story lessons were longer on average than the comparison-group lessons. Our impression throughout the conduct of this study was that SAIL students take more

time when reading orally, with teachers frequently interjecting explicit explanations, requesting think-alouds, and elaborating responsively. Thus, we believe that at least the increased interpretations in the SAIL condition were due more to how time was spent in the SAIL lesson than to amount of time per se, although the design of this study does not permit a definitive conclusion on this point.

Spring think-aloud analysis

The think-aloud protocols generated by each student in reaction to the Aesop's fable about the dog and his reflection were transcribed and analyzed using an analytic induction approach (Goetz & LeCompte, 1984). Two raters (one rater was "blind" to condition) read through all of the protocols, independently taking notes and identifying potential categories of reported reading processes (see Footnote 4). Through negotiation, a tentative set of process categories were identified, and these were applied by both raters independently to two protocols, one from a SAIL student and one from a comparison-group student. The two raters then met and refined the categories in light of the difficulties experienced scoring these two protocols. The refined categorization was applied to another pair of protocols, again independently by both raters. The refined categorizations captured all of the processes represented in these protocols, and thus, this set of processes was used to code all of the think-aloud protocols.

A response with any indication of comprehension strategy use was coded as "strategy-based." For example, the following excerpt was coded as a strategies-based response:

(The student read the page about the dog rushing out of the house with the piece of meat. The student then started to talk before the researcher asked an initial probe.)

S: I think my prediction is coming out right. (verifying)
R: Why do you say that?
S: Cuz, cuz I see a bridge over there and water. (using picture clues)
R: Uh huh. . . .
S: And he ran out of the house without anybody seeing him. Like I said before. . . .
R: Okay, so you think your prediction is right and you're using, you were pointing to the pictures.
S: Yep.

The specific strategies used were also coded using the comprehension strategy definitions from the strategies interview, with 89% agreement between

two raters on 20% of the protocols on these codings of specific strategies. The mean number of strategies evidenced by SAIL reading group members (averaging across all groups) was 6.93 (SD = 1.46). The corresponding comparison-group mean was 3.18 (SD = 1.06). The SAIL readers applied significantly more strategies during the think-aloud task than did the comparison-group students, $t(4) = 9.59$, $p < .05$. In fact, there was no overlap in the group means, with SAIL group means ranging from 5.00 to 8.67 strategies used per student, on average, and corresponding comparison-group means ranging from 2.00 to 4.83. All strategies that were scored, except for one (monitoring), were observed descriptively more frequently in the SAIL than in the comparison protocols. The strategies that occurred in the SAIL condition, from most to least frequent, were as follows: predicting, relating text to prior knowledge, thinking aloud, summarizing, using picture clues, verifying, seeking clarification, monitoring, looking back, visualizing, and setting a goal. The corresponding order for the comparison condition was predicting, using picture clues, verifying, relating text to prior knowledge, monitoring, seeking clarification, thinking aloud, and looking back. No apparent visualizing, summarizing, and setting a goal were observed in the comparison-group think-alouds.

We also examined whether SAIL or comparison groups focused more on text- or reader-based information when they did not respond strategically. Responses not classified as strategies-based were coded as either "text-based" or "reader-based" (interrater agreement on 20% of the protocols for classifying text- or reader-based responses was 94%).

Text-based responses contained information explicitly stated or pictured in the story. For example, after reading the first text segment, a student responded to the initial probe:

R: Okay, what are you thinking?
S: The dog stole something.
R: Uh huh . . . tell me more.
S: He knocked over the table.
R: He knocked over, talk nice and loud . . . he knocked things off the table . . . okay.
S: Yeah, and nothing really else.
R: Okay. And what do you think about what the dog did?
S: What do you mean?
R: What do you think about what the dog did?
S: He stole something.

Reader-based responses reflected a connection between the story and a student's prior knowledge, experiences, beliefs, or feelings. In the following example, a student read the segment about the dog stealing a piece of meat from the master's dinner table:

R: What are you thinking about what's happening on this page?
S: Sort of bad because I see that was part of their dinner, but they would not have all the uhm, protein.
R: Okay. . . .
S: The dog ate all that. . . .

Proportions were calculated for each class, indicating the relationship of text- and reader-based responses to the total number of responses that were not coded as strategies-based. From these class proportions, SAIL and comparison group means were computed. The mean for reader-based responding for the SAIL group was .74 (SD = .10). The mean proportion of reader-based responding for the comparison group was .45 (SD = 0.17). Thus, the SAIL group produced more reader-based responses than the comparison group, $t(4) = 3.61$, $p < .05$. Without exception, all SAIL classes were proportionally more interpretive than literal in their nonstrategies-based responses. In contrast, only 2 of 5 comparison classes were proportionally more interpretive in their responses.

In summary, the SAIL students used strategies on their own more than the comparison students. Although strategy use by itself does not constitute self-regulation, it does suggest that students had begun to apply strategies independently, one aspect of self-regulated reading. Self-regulated readers are not only strategic; they also are goal-oriented, planful, and good comprehension monitors. Because we did not ask students to report directly on their strategic processing while reading, however, we cannot address those aspects.

In addition, the results of the think-aloud analysis supported the results of the recall analyses. For a story in which variable instructional time was not a factor, SAIL students made significantly more reader-based remarks than comparison students. The SAIL students responded more interpretively as well as personally.

Spring standardized test performance

In May–June, the SAIL students outperformed the comparison students on the 40-item comprehension subtest. The reading group raw score mean in the SAIL condition was 34.20 (SD = 2.65); the corresponding comparison-group mean was 28.73 (SD = 3.77), $t(4) = 4.02$ (see Table 1). The SAIL students also outperformed the comparison students on the 36-item word skills subtest, $t(4) = 3.98$. The reading group word skills raw score mean in the SAIL condition in the spring was 27.10 (SD = 2.19); the corresponding comparison-group mean was 24.00 (SD = 1.53).

One of the most striking aspects of the spring comprehension standardized test data was the much lower variability among individual students within SAIL reading groups compared to comparison reading groups. (The

careful matching of the reading groups in the fall was with respect to both mean performance *and* variability on standardized reading comprehension; thus, there was little difference in SAIL and comparison reading group variabilities in the fall, as reported in the Method section.) Also, with the exception of one pair of classes (T5 and T10), this lower variability among students in SAIL reading groups was evident in the spring word study skills data. This finding is obvious from examination of the standard deviations for each matched pair of reading groups on the standardized subtests (see Table 1).

We believed that an especially strong demonstration of the efficacy of the SAIL program would be greater gains on standardized measures over the course of the academic year in SAIL versus the comparison condition. Thus, we tested the size of the fall-to-spring increase in raw scores in the SAIL versus the comparison groups. The SAIL group averaged 22.20 on an alternate form of the comprehension subtest (SD = 6.85) at the late fall testing, indicating a fall-to-spring gain of 12.00 (SD = 5.20) on average, and the comparison classes averaged 22.67 (SD = 5.89) in the fall, yielding a fall-to-spring change of 6.07 (SD = 2.28) on average. For the word skills subtest, the fall SAIL mean was 20.97 (SD = 2.76), and the mean fall-to-spring increase was 6.13 (SD = 1.86). In the fall, the comparison mean was 21.10 (SD = 3.40), and the fall-to-spring mean difference was 2.90, (SD = 2.70). The one-tailed interaction hypothesis test was significant, as anticipated for the comprehension subtest, $t(4) = 3.70$. The word skills subtest proved significant as well, $t(4) = 5.41$.

In one of the matched pairs, there were some perfect scores on the comprehension posttest: The SAIL class mean was 36.83 (SD = 2.40); the non-SAIL class mean was 35.17, SD = 4.22). For this pair of reading groups, a version of the next level of the Stanford Comprehension subtest (Primary 2, Form J) was then administered. Consistent with the analyses reported in the last two paragraphs, the spring SAIL group mean was greater than the matched comparison-group mean, and the SAIL group standard deviation was lower than the comparison-group standard deviation: SAIL M = 29.8, SD = 5.42; comparison-group M = 21.8, SD = 10.17. (The pretest Reading Comprehension subtest mean for the SAIL class was 33.83 [SD = 7.28]; the mean for the non-SAIL class was 32.17 [SD = 6.88]).

In summary, by academic year's end, the SAIL second-grade students clearly outperformed the comparison-group students on standardized tests, with greater improvement on the standardized measures over the course of the academic year in the SAIL condition. Unfortunately, no additional end-of-year achievement data existed for the students for comparison purposes either in reading or in any other subject area.

On the standardized tests, gains in comprehension were expected because, more than anything else, SAIL is intended to increase students' understanding of text. The effects on students' word skills performance were more of a surprise, albeit a pleasant one, supportive of the SAIL intervention; we

knew that all teachers, regardless of condition, taught phonics and word attack skills, although at different times of day (e.g., integrated into various content areas) and in different ways (e.g., covered in the form of worksheets or mini-lessons).

Discussion

We made many informal and formal observations throughout the 1991–1992 school year indicating that instruction in the SAIL and comparison classes was very different. The differences were apparent in the two lessons that were analyzed in the spring: A SAIL curriculum developer and several graduate students who were familiar with transactional strategies instruction had no difficulty discriminating between transcripts of SAIL and non-SAIL lessons. One important difference highlighted in the analysis of the spring lessons was that discussion of strategies was much more prominent in the SAIL than in the comparison reading groups. That the differences in teaching were so clear bolsters our confidence in this study as a valid assessment of the efficacy of SAIL with at-risk second-grade children.

SAIL had positive short-term and long-term impacts. In the short term, students acquired more information from stories read in reading group and developed a richer, more personalized understanding of the stories. Whether the focus is on the amount of literal information recalled from stories covered in reading group or student interpretations of the texts read, there were indications in these data of superior performance by SAIL students relative to the comparison students. We infer that SAIL students learn more daily from their reading group lessons than do students receiving more conventional, second-grade reading instruction.

SAIL had long-term impacts as well. Consistent with our expectations, the SAIL students exhibited greater awareness of strategies by the end of the year than the comparison students. SAIL students also reported use of, or were inferred to use, strategies more than the comparison students: They thought aloud while reading the Aesop's fable at the end of the year. The standardized test performances of the SAIL students also were superior to those of the comparison students at the end of the year. Most critically, there was significantly greater improvement on standardized measures of reading comprehension from fall to spring in the SAIL versus the comparison classrooms. In short, all measurements of student reading achievement reported here converged on the conclusion that a year of SAIL instruction improves the reading of at-risk second-grade students more than does alternative high quality reading instruction.

This study is the strongest formal evidence to date that transactional strategies instruction improves the reading of elementary-level students. There were many elements taken into consideration in this study that varied freely in more informal comparisons of SAIL and alternative instruction, such as

ones generated by the school district that developed the intervention: (a) The student participants were carefully matched in this investigation so that there was no striking difference in their standardized reading achievement at the outset of the study. (b) The teachers were carefully selected. From years of observing and interviewing committed SAIL teachers, we knew that they are excellent teachers in general, who offer rich language arts experiences for their students. Thus, it was imperative that a compelling evaluation of SAIL be in comparison to excellent second-grade instruction. Accordingly, we sought highly regarded comparison teachers. (c) The lessons analyzed in the transactional strategies instruction and comparison groups involved the groups' processing the same stories. (d) The same dependent measures were administered by the same tester so that measurement experiences were equivalent for participants.

Another strength of this evaluation was that it did not rely only on standardized assessments but included also assessments of students' reading that were grounded in their typical classroom experiences. The assessments of children's memories for and interpretations of stories read in class reflect better the day-to-day comprehension demands on students than do standardized measures. Although thinking-aloud measures are far from perfect indicators of thinking (Ericsson & Simon, 1980), the assessments of children's thinking as they read the Aesop's fable arguably tapped more directly the thinking processes of the children that SAIL was intended to change than did the standardized assessments.

Are the outcomes reported here generally significant beyond the specifics of the SAIL program? SAIL is a specific instantiation of reading comprehension strategies instruction as adapted by educators. Such instruction may serve as a model for other educators. SAIL provides teachers with a way to blend critical elements of direct teaching and holistic principles of instruction, aspects of instruction that may already exist in conventional reading classrooms. Because many conventional programs already share features with SAIL (e.g., literature-based instruction, teaching of predicting and problem-solving strategies), these programs might be modified to include SAIL components.

As we argued at the beginning of this article, long-term, direct explanation of thinking processes and scaffolded practice of a manageable repertoire of powerful comprehension strategies constitute an approach replicated in a number of settings (see also Pressley, El-Dinary, et al., 1992, and Pressley & El-Dinary, 1993, for a number of examples). The practice has raced ahead of the science, however, with the educator-developed adaptations more ambitious in scope, more complex, and ultimately very different from the researcher-validated interventions (e.g., reciprocal teaching) that inspired the educator efforts. There is a real need to evaluate such adaptations, for there is no guarantee that the strategies instruction validated in basic research studies is effective once it is translated and transformed dramatically by educators.

The research reported here contrasts with basic research on strategies instruction in a number of ways. First, the intervention studied here was multicomponential and this study was not analytical at all with respect to components of the intervention. Typically, basic strategies instruction research has been much more analytical. We can defend this evaluation of an entire transactional strategies instruction package because the whole program is the unit of instruction in the schools we have been studying: When the interest is in whether an instructional package as a whole works, a study evaluating that whole relative to other instruction is definitely defensible, particularly if time spent in direct instructional activities is controlled carefully (e.g., in this study, both groups of students received a year of reading instruction in the context of a full year in the second grade). Moreover, it was not our intent to tease out which aspects of the program were most effective nor to determine which components in combination accounted for student gains, especially because we believe that the complex instruction exemplified by SAIL may be more than the sum of its component parts (Pressley, El-Dinary, et al., 1992).

Second, the program of research that includes this study is a mix of qualitative and quantitative research. In contrast, most basic studies of strategies have been quantitative only. We are certain that the quantitative study reported here would have been impossible without the 3 years of qualitative research leading up to it. At a minimum, that qualitative research affected the selection of dependent measures and the decision to study only accomplished SAIL teachers (see Pressley, Schuder, et al., 1992). More generally, it made obvious to us the scope of an investigation necessary to evaluate transactional strategies instruction so that the treatment would not be compromised by the evaluation.

Third, most basic strategies research is designed and conducted by researchers. When educators have participated in basic studies, it has been as delivery agents only. In the program of transactional strategies instruction research, researchers, program developers, and educators have combined their talents to produce a body of research that realistically depicts transactional strategies instruction and evaluates it fairly. As the study was designed and as it unfolded, school-based educators were consulted frequently about the appropriateness of potential dependent measures and operations of the study. The result has been a much more complete and compelling set of descriptions of transactional strategies instruction and, now, a thorough appraisal of the impact of one transactional strategies instruction program on second-grade, weaker readers.

We do not claim that after 1 year of transactional strategies instruction these students have become self-regulated readers. Pressley and Afflerbach (1995) made the point that truly self-regulated reading is observed only in very mature readers. It has always been suggested that TSI needs to occur over the long term to be effective (Pressley, El-Dinary, et al., 1992). Our

hypothesis is that true self-regulation is the product of years of literacy experiences, with TSI intended to get the process off to a good start. One year of such instruction at least gets second-grade readers who are experiencing difficulties in learning to read to improve their reading relative to a year for comparable students in very good conventional classrooms.

Acknowledgments

The work reported in this article was supported in part by the Educational Research and Development Center Program, National Reading Research Center (NRRC), University of Maryland; PR/AWARD NUMBER 117A20007, as administered by the Office of Educational Research and Improvement (OERI), U.S. Department of Education. The findings and opinions do not necessarily reflect the opinions or policies of NRRC, the OERI, or the U.S. Department of Education. We are grateful for the input of a number of University of Maryland and school-based collaborators, including Pamela B. El-Dinary, Jan Bergman, Laura Barden, Marsha York, and the 10 teachers who so graciously permitted this research to be conducted in their classrooms during 1991–1992.

Notes

1 We recognize that some readers may be concerned about the mean difference in years of teaching between the SAIL and non-SAIL teachers. In this study, the SAIL and non-SAIL classes were matched as closely as possible. The primary criteria for matching classes were demographic in nature. To the extent that it was possible, we used student mobility patterns, Chapter 1 status, ethnic and minority composition, size and location of schools, and standardized test performances. At the time, years of teaching experience did not seem to be as critical as some of the other factors. Given our decision, there is no way to separate out the effect that years of experience may have had on the way teachers taught their students. However, readers should bear in mind that the comparison teachers were highly regarded for their teaching abilities by district personnel; therefore, if anything, their greater number of years of experience could be construed as an advantage.

2 All class means were based on 6 students with the exception of the following, which reduced this number because of either data loss or absence: Strategies interviews: 1 student in two non-SAIL classes (pretest), 1 student in one SAIL class (posttest); retellings: 1 student in one SAIL class ("Mushroom" story), 1 student in one SAIL class ("Fox Trot" story); think-aloud task: 2 students in one SAIL class.

3 One reviewer strongly felt that the skipping strategy was not as "good" or "useful" as some of the other strategies students reported using. Consequently, we are providing data so that readers can compare the two groups specifically on the skipping strategy. For the fall strategies interview, the SAIL sum of mean frequencies by group for skipping as a word attack strategy was 3.93 ($SD = 0.35$), and the non-SAIL summed mean was .94 ($SD = 0.70$), $t(4) = 8.20$, $p < .05$. For the spring

strategies interview, the SAIL summed group mean for skipping as a word attack strategy was 4.33 (SD = 1.10), and the non-SAIL summed mean was 1.47 (SD = 1.55), $t(4) = 5.36$, $p < .05$. The interaction was $t(4) = -0.36$, $p > .05$. For the fall strategies interview, the SAIL sum of mean frequencies by group for skipping as a *comprehension strategy* (i.e., ignoring a larger segment of text and reading on) was .17 (SD = 0.40), and the non-SAIL summed mean was .57 (SD = 0.50), $t(4) = -1.10$, $p > .05$. For the spring strategies interview, the SAIL summed group mean frequency was 2.47 (SD = 1.25) and the non-SAIL mean frequency was .83 (SD = 0.80), $t(4) = 2.14$, $p < .05$. The interaction was also significant, $t(4) = 2.20$. (The permutation test for the interaction was not significant, however, because of one tie in the data.)

4 We recognize that to rule out possible alternative explanations of the results, the two raters conducting interrater agreement should be "blind." However, there is a perspective held by some qualitative researchers that the use of blind raters does not do justice to the analysis of data because the blind rater has spent so little time immersed in the experiences that have led to the primary researcher's breadth of understanding. Thus, "expecting another investigator to have the same insight from a limited data base is unrealistic" (Morse, 1994, p. 231). We concurred to some extent with this argument; however, in attempting to reconcile positions, we opted for only one rater to be blind. In that way, the blind rater could lend credibility to the nonblind researcher's interpretations. In attempting to strike a balance, the nonblind researcher often deferred to the blind rater's opinion when a stalemate was reached. Also, when the primary researcher was unsure how to interpret the data in the transcripts and protocols that were not subjected to interrater agreement, the "blind" rater frequently assisted in the coding of the questionable segment or unit.

References

Allington, R. L. (1991). The legacy of "Slow it down and make it concrete." In J. Zutell & S. McCormick (Eds.), *Learner factors/teacher factors: Issues in literacy research and instruction: Fortieth yearbook of the National Reading Conference* (pp. 19–29). Chicago: National Reading Conference.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353–394). New York: Longman.

Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review*, *75*, 81–95.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction*, *2*, 131–156.

Bergman, J., & Schuder, R. T. (1992). Teaching at-risk elementary school students to read strategically. *Educational Leadership*, *50*(4), 19–23.

Brown, R. (1995a). *A quasi-experimental validation study of strategies-based instruction for low-achieving, primary-level readers*. Unpublished doctoral dissertation, University of Maryland at College Park.

Brown, R. (1995b, April). *The teaching practices of strategies-based and non-strategies-based teachers of reading*. Paper presented at the meeting of the American Education Research Association, San Francisco, CA.

Brown, R., & Coy-Ogan, L. (1993). The evolution of transactional strategies instruction in one teacher's classroom. *Elementary School Journal*, *94*, 221–233.

Brown, R., & Pressley, M. (1994). Self-regulated reading and getting meaning from text: The transactional strategies instruction model and its ongoing evaluation. In D. Schunk & B. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 155–179). Hillsdale, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, C. (1991). Reading instruction that increases thinking abilities. *Journal of Reading*, *34*, 510–516.

DeFord, D. (1985). Theoretical orientation to reading instruction. *Reading Research Quarterly*, *20*, 351–367.

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, *61*, 239–264.

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, *22*, 347–368.

Durkin, D. (1978–1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, *12*, 481–538.

Eeds, M., & Wells, D. (1989). Grand conversations: An exploration of meaning construction in literature study groups. *Research in the Teaching of English*, *23*, 4–29.

El-Dinary, P. B., & Schuder, R. T. (1993). Seven teachers' acceptance of transactional strategies instruction during their first year using it. *Elementary School Journal*, *94*, 207–219.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.

Garner, R. (1988). Verbal report data on cognitive and metacognitive strategies. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 63–76). San Diego, CA: Academic Press.

Gaskins, I. W., Anderson, R. C., Pressley, M., Cunicelli, E. A., & Satlow, E. (1993). Six teachers' dialogue during cognitive process instruction. *Elementary School Journal*, *93*, 277–304.

Ginsburg, M. (1991). Mushroom in the rain. In D. Alvermann, C. A. Bridge, B. A. Schmidt, L. W. Searfoss, P. Winograd, & S. G. Paris (Eds.), *My best bear hug* (pp. 144–154). Lexington, MA: Heath.

Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research*. San Diego, CA: Academic Press.

Golden, J. M. (1988). The construction of a literary text in a story-reading lesson. In J. L. Green & J. O. Harker (Eds.), *Multiple perspective analyses of classroom discourse* (pp. 71–106). Norwood, NJ: Ablex.

Harris, A. J., & Sipay, E. R. (1985). *How to increase reading ability: A guide to developmental and remedial methods*. New York: Longman.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (NCTE Research Report No. 3). Champaign, IL: National Council of the Teachers of English.

Hutchins, E. (1991). The social organization of distributed cognition. In L. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 283–307). Washington, DC: American Psychological Association.

Johnston, P., & Afflerbach, P. (1985). The process of constructing main ideas from text. *Cognition and Instruction*, *2*, 207–232.

Kirk, R. E. (1982). *Experimental design* (2nd ed.). Monterey, CA: Brooks/Cole.

Lytle, S. L. (1982). *Exploring comprehension style: A study of twelfth-grade readers' transactions with text*. Unpublished doctoral dissertation, University of Pennsylvania. (University Microfilms No. 82-27292).

Marks, M. B., Pressley, M., Coley, J. D., Craig, S., Gardner, R., DePinto, T., & Rose, W. (1993). Three teachers' adaptations of reciprocal teaching in comparison to traditional reciprocal teaching. *Elementary School Journal*, *94*, 267–283.

Marshall, E. (1982). *Fox in love*. New York: Penguin Books.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

Meichenbaum, D. (1977). *Cognitive-behavior modification: An integrative approach*. New York: Plenum Press.

Miller, J. P. (1976). *Tales from Aesop*. New York: Random House.

Morse, J. M. (1994). Designing funded qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 220–235). Thousand Oaks, CA: Sage Publications.

O'Flahavan, J. F. (1989). *Second graders' social, intellectual, and affective development in varied group discussions about literature: An exploration of participation structure*. Unpublished doctoral dissertation, University of Illinois at Champaign-Urbana.

Olshavsky, J. E. (1976–1977). Reading as problem-solving: An investigation of strategies. *Reading Research Quarterly*, *12*, 654–674.

Olson, G. M., Mack, R. L., & Duffy, S. A. (1981). Cognitive aspects of genre. *Poetics*, *10*, 283–315.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Paris, S., & Oka, E. R. (1986). Children's reading strategies, metacognition, and motivation. *Developmental Review*, *6*, 25–56.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Pressley, M., & El-Dinary, P. B. (Guest Eds.). (1993). Special issue on strategy instruction. *Elementary School Journal*, *94*(2).

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *Elementary School Journal*, *92*, 513–555.

Pressley, M., Gaskins, I. W., Cunicelli, E. A., Burdick, N. J., Schaub-Matt, M., Lee, D. S., & Powell, N. (1991). Strategy instruction at Benchmark School: A faculty interview study. *Learning Disability Quarterly*, *14*, 19–48.

Pressley, M., Goodchild, R., Fleet, J., Zajchowski, R., & Evans, E. D. (1989). The challenges of classroom strategy instruction. *Elementary School Journal*, *89*, 301–342.

Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J. A., & Kurita, J. A. (1989). Strategies that improve children's memory and comprehension of text. *Elementary School Journal*, *90*, 3–32.

Pressley, M., Schuder, T., Teachers in the Students Achieving Independent Learning Program, Bergman, J. L., & El-Dinary, P. B. (1992). A researcher–educator collaborative interview study of transactional comprehension strategies instruction. *Journal of Educational Psychology*, *84*, 231–246.

The Psychological Corporation (1990). *Stanford Achievement Test Series: Technical data report* (8th ed.). San Diego, CA: Harcourt Brace Jovanovich.

Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactional theory of literary work*. Carbondale: Southern Illinois University Press.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*, Educational Research Association, Chicago, IL. 479–530.

Schuder, R. T. (1993). The genesis of transactional strategies instruction in a reading program for at-risk students. *Elementary School Journal*, *94*, 183–200.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scriber, & E. Souberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.

Wyatt, D., Pressley, M., El-Dinary, P. B., Stein, S., Evans, P., & Brown, R. (1993). Comprehension strategies, worth and credibility monitoring, and evaluations: Cold and hot cognition when experts read professional articles that are important to them. *Learning and Individual Differences*, *5*, 49–72.

Appendix: Summary of "mushroom in the rain" lessons

SAIL Teachers

Teacher 1

The teacher reviewed what expert readers do. She questioned students about the strategies good readers apply when reading. She augmented their responses, explaining some benefits of strategies use. She reviewed with students what they could do when they came to an unknown word (e.g., use picture clues, guess, skip, look back in text). She also focused on verbalizing thinking, summarizing, and visualizing. She asked students to browse through pages and make predictions. A student predicted that the story might be like "The Mitten," a story the group had read earlier in the year. Students discussed possible connections between the two stories. The teacher directed students to verify their predictions as they read and had them visualize a descriptive segment. Students took turns reading. When they finished reading, students either thought aloud spontaneously or were cued to do so by the teacher. Thinking aloud consisted of summarizing content, voicing an opinion, suggesting an interpretation, making or refining predictions, or relating text content to background knowledge or personal experiences. After the

reader thought aloud, other students were encouraged to elaborate, persuade, or counter the interpretation.

Students frequently supported their interpretations with background knowledge or text clues. Students continued to discuss similarities and differences between "Mushroom in the Rain" and "The Mitten." For example, they debated whether the mushroom was growing or stretching. Students practiced sequencing by summarizing story content. During discussions, the teacher restated students' responses, clarified confusions, sought elaborations, and garnered opinions from group members. When students faced a word they did not know, they were urged to use one of their "fix-up" strategies. The teacher generally did not ask specific questions about text details. At the end of the lesson, students verified their predictions and fine-tuned their interpretations using text information and background knowledge. Several students admitted they were confused by aspects of the story. When the teacher asked what they could do about this, a student suggested they reread the story. The teacher replied that a good strategy to clarify confusions was rereading. The lesson ended with a student summarizing the story.

Teacher 2

The teacher reviewed what good readers do. Students described the various strategies and evaluated their usefulness. When students talked about visualizing, the teacher explained a personal use of the strategy. The teacher discussed with students the flexible application of a coordinated set of strategies. She encouraged students to use their strategies during story reading. The teacher told students she would focus on visualizing in the lesson. She read the title and first page, modeling her thinking as she visualized text content and made connections between the story and her experiences. She encouraged students to relate the story to their own experiences. Without prompting, a student predicted that the story would be like "The Mitten," a story the class read earlier in the year. The teacher asked the student to support his claim. Students took turns reading aloud. When they came to an unknown word, they often used strategies without teacher prompting. When they needed help, the teacher cued them to use one of their problem-solving strategies ("fix-it kit"). After reading a page, students would think aloud on their own or be prompted to do so by the teacher. When thinking aloud, students summarized story content, made predictions, or offered interpretations. Other students would then respond to the first student's remarks.

Students continually discussed how "Mushroom in the Rain" was similar to "The Mitten." The group referred to different versions of the story. At one point, a student observed that the animals going under the mushroom were increasing in size. When observations like this one were given by students, the teacher told the group to bear them in mind as they read. Students made and verified predictions frequently and related events to their background

knowledge and personal experiences. They elaborated on each other's ideas. During discussions, the teacher did not state her own opinion. Instead, she rephrased students' comments or sought elaboration. When the group thought about what happens to a mushroom in the rain, some students believed the mushroom grew; others countered that the animals stretched it. The teacher allowed students to choose the interpretation they favored. The teacher praised students for their use of strategies, such as making connections between "The Mitten" and "Mushroom in the Rain." She encouraged the group to continue to use strategies in future years because they would help them become better readers.

Comparison Teachers

Teacher 7

The teacher reviewed new words that were presented on cards in the context of sentences. Students were prompted to use the word attack strategies they had been practicing: looking at the first sound, proceeding to the vowel, and then seeing if the word had a suffix. Students took turns reading the story aloud. When students had difficulty, the teacher prompted them to use their word attack strategies and sometimes she gave them the word. After students read, the teacher periodically summarized what had transpired. She drew students' attention to the illustrations. She asked students literal and interpretive comprehension questions about the text, activated their background knowledge, solicited their opinions, and allowed divergence in interpretations (e.g., "Does the ant want to share the mushroom? What does the mushroom remind you of? What do you use in the rain?"). These questions typically did not generate extended discussion. When a student mentioned that the butterfly couldn't fly because his wings were wet, the teacher reminded students of their unit on butterflies. One topic students had been learning about was "persuasion"; the teacher related this topic to the way the animals were persuading the ant to let them under the mushroom. After reading a section, the teacher often asked students what they were thinking. The teacher taught new vocabulary in context, relating word meanings to students' background knowledge.

At one point, the teacher drew a mushroom on the board. She asked students to tell her the order of animals that went under the mushroom. She questioned how all the animals fit under the mushroom. She related this story to other stories students had read. One student said the mushroom grew because of the rain. She confirmed that mushrooms grow rapidly in the rain. When students faced unfamiliar words, she directed them to apply their word attack strategies and knowledge of phonics (e.g., "Good boy, it's got that double *p* to keep that *o* short. . . ."). After reading one section, she drew students' attention to the quotation marks, colon, commas, and

exclamation mark that were on the page. She asked for predictions, without requesting support for students' ideas. Some interpretive discussion occurred around the nature and motives of the fox. When adding the fox to the sequencing on the chalkboard, she said, "When you're making a sequence and you're writing a story or reading it, sometimes it's nice to make an illustration, and then you can add words underneath it to help you organize, get things in, what happened first, second, third, next, and then final." After reading, the teacher frequently drilled students on word skills, using words from the story. Students received a "point" for answering questions correctly. She asked students to find words with suffixes and base words. She frequently provided direct instruction of rules (e.g., making plurals from singular forms; "To keep the *i* short before you add a suffix that begins with a vowel like *-er, -ing, -est, . . . -ious*, we have to make sure there's two consonants, to double the letter."). Periodically, she complimented students on their thinking. After reading, students pretended to touch a mushroom. She asked for descriptive words and similes. At the end of the lesson, the teacher told students to visualize to help them remember the ordering of story events. She informed students that they would retell and illustrate the story the next day.

Teacher 10

The teacher stated the title of the story. She asked the students to read the first three pages silently, looking for words they did not know. As students pointed out unfamiliar words, the teacher helped them with word clues. For example, she said that "one of the ways we can find out what a word is sometimes, if we're not too sure of it, is to see if there are little tiny word clues inside of a big word and that will help sound out the word. . . . That's a good word attack skill." The teacher then had a student read the first page. She directed the group to look at the illustration. She told them to notice the size of the mushroom and to watch out for what happens. There was little discussion during story reading. However, at one point, a student volunteered that the story was like "The Mitten." The teacher did not elaborate on the student's comment except to say "let's see what happens." Toward the end of the story, the teacher asked what happened to the mushroom. One student said the mushroom grew. When the teacher asked why, he answered that it was because, "the water came in the soil and made it grow." At the end of the story, the teacher said that the student "found the secret. That was the secret of how they all fit." Others concurred. One student pointed to the picture of the mushroom getting bigger. The teacher elaborated. "All right, so S found out because he was watching the pictures and getting a clue from the pictures." The group talked a little more about plants needing lots of water to grow. The teacher asked students to tell about any character they liked and what they liked about him. Several students gave opinions.

Discussion then centered on the fox's nature. Students used their prior knowledge to state that the fox was smart. The teacher redirected students to a specific page, asking them to look for a clue. The students recognized that the fox was tricked, and they changed their minds. The group spent much time discussing this episode and looking at the picture. The teacher asked students to fold a piece of paper into four sections. She asked them to draw in order what happened to the mushroom, telling them they could refer to the book for help. She guided them through the activity. The teacher then asked students to suggest alternate endings. Several students responded. She asked students to web the character traits of one of the animals. She told them to "go back into your story and see if there are any story clues . . . and think of some words that would describe that particular character." Students took turns sharing their webs and descriptive words. The teacher asked if they liked the story and whether "it had a nice moral to it. Was it a good lesson about kindness?" Students assented but did not discuss their reasons. She suggested students put new words they learned in their ABC books (i.e., personal word books).

68

UNDERSTANDING READING COMPREHENSION

Current and future contributions of cognitive science

R. F. Lorch Jr. and P. van den Broek

The ability to read and comprehend text is crucial for success in our society and its development has been a main component of instructional practice. In the past 2 decades, psychologists have devoted a good deal of attention to the question of how competent, adult readers comprehend text. Influenced by work in linguistics and artificial intelligence, the efforts of these cognitive scientists have dramatically increased our understanding of the psychological mechanisms underlying reading comprehension. In this article, we provide an overview of the contributions of cognitive research on text comprehension and an agenda for future research. Our specific interest is in understanding the processes by which skilled adult readers comprehend text.

We first sketch the historical progression of experimental research on text processing and then present more detailed analyses of the major contributions of cognitive science. The emphasis is on the theoretical developments that have played central roles in advancing our understanding of text comprehension. Our premise is that by identifying how progress has been achieved in understanding central aspects of text comprehension, we will be in better position to suggest how research might best proceed in domains where less progress has occurred. Finally, we identify what we consider the most pressing topics for future research and suggest the kinds of theoretical developments that appear necessary to advance research in those domains.

Historical context

Early research on text comprehension in experimental psychology

Bartlett's (1932) study is widely credited as the first serious investigation of text comprehension and memory in experimental psychology (e.g., Lachman,

Source: *Contemporary Educational Psychology*, 1997, 22, 213–246.

Lachman, & Butterfield, 1979). Bartlett observed that readers' memories for textual information were systematically distorted to fit their own factual and cultural knowledge and that the distortion increased with time. Thus, he demonstrated that a reader's understanding and memory for text is an active, constructive process rather than a passive, receptive process. Several subsequent investigations of text processing fall squarely within the theoretical orientation he established. Bartlett distinguished between a text's surface representation and the form of the mental representation constructed by the reader. This distinction was supported by early research on text processing (Bransford, Barclay, & Franks, 1972; Bransford & Franks, 1971; Sachs, 1967). Bartlett also emphasized that readers attempt to construct an understanding of a text that is coherent both in the sense of having internal organization and in the sense of being interpreted with respect to the reader's prior knowledge. Consistent with this position, several early studies demonstrated that the theme of a narrative is central to a reader's understanding of a text (Bransford & Johnson, 1972; Pompi & Lachman, 1967; Sulin & Dooling, 1974) and that memory for a text depends critically on a reader's ability to relate text content to appropriate background knowledge (Bransford & Johnson, 1972; Dooling & Lachman, 1971). Also relevant in this regard is research in educational psychology demonstrating that appropriate advance organizers help readers establish coherence and thereby facilitate memory for text content (Ausubel, 1960).

Distinct from the schema-theoretic approach established by Bartlett are two related literatures with neobehaviorist origins (see Anderson & Bower, 1973). Both lines of research were conducted primarily by educational psychologists interested in applying basic experimental research on memory to classroom learning. One line of research examined whether the interference theory of forgetting that was developed to explain paired associate learning also provided an adequate account of forgetting of text. In fact, when care is taken to establish appropriate text and testing conditions, interference effects are found with connected discourse (Anderson & Myrow, 1971; Crouse, 1971; Myrow & Anderson, 1972). The conditions under which interference is observed are relatively restricted, however, so the theory lacks sufficiency as an account of text memory.

The other line of research examined the influence on text memory of questions inserted in the text. This research demonstrated that adjunct questions facilitate memory for text content in systematic ways. Memory for content specifically relevant to the adjunct questions is improved regardless of whether the questions precede or follow the targeted content in the text (Frase, 1968, 1969; Rothkopf, 1966). However, the memory benefits of adjunct questions are restricted to the targeted content when the questions precede the relevant content, whereas more general facilitative effects result when the adjunct questions follow the text sections containing the targeted information (Rothkopf, 1966; Rothkopf & Bisbicos, 1967).

Despite its historical importance, the research on text processing conducted prior to 1973 is curiously dissociated from research after that point in time. There are several probable reasons for this discontinuity. First, the research literatures on advance organizers, adjunct questions, and interference effects on text memory were mainly concerned with the implications of empirical studies for classroom learning. The construction of a cognitive theory of text comprehension was not a primary goal of this research. Second, although more directly concerned with the nature of text comprehension, research following the tradition of Bartlett (e.g., Pompi & Lachman, 1967; Bransford & Franks, 1971) lacked a clear theoretical model that might have sustained the research and given it direction. Third, none of the research efforts were based on a well-developed conception of how to represent information in a text or in a reader's memory. Finally, when this research sought to make inferences about the nature of comprehension processes occurring during reading, it relied almost exclusively on the use of memory measures taken after reading. Although, at the time, there was some recognition of the perils of using "off-line" memory measures to study "on-line" comprehension processes (ef. Carroll, 1972), research since the early 1970s has made it abundantly clear that memory measures taken after reading are all too indirect as indicators of processes occurring during reading.

Building a foundation for research on text comprehension processes

Text comprehension did not really come into its own as a domain of inquiry within experimental psychology until the 1970s. Several developments fueled interest in the area and set the stage for an explosion of research on text comprehension in the 1980's. In this section, we note what we consider the four most significant developments of the 1970's.

First, between 1973 and 1976, several cognitive scientists published ambitious theories addressing a broad range of issues concerning the encoding, representation, retrieval, and application of linguistic (and other types of) knowledge. These theories included HAM (Anderson & Bower, 1973). ACT (Anderson, 1976), ELINOR (Norman & Rumelhart, 1975), spreading activation theory (Collins & Loftus, 1975; Collins & Quillian, 1969), conceptual dependency theory (Schank, 1975), and the theories of Kintsch (1974) and Miller and Johnson-Laird (1976). Of these, only Kintsch's research directly addressed the topic of text comprehension, but each theory was sufficiently broad to have relevance for text comprehension. Of interest in the current context are the points of consensus of the models with respect to the representation of complex information in memory (Lachman *et al*., 1979). All of the models were concerned with developing a system that was sufficient to represent complex knowledge acquired through both verbal and nonverbal interactions with the world. All of the models sought a representational system that allowed efficient search and retrieval. And all of the models attempted

to incorporate a representational system that allowed for rapid inference. The solution chosen by each of the theoretical enterprises was to adopt the *proposition* as a basic unit of meaning and to represent relations among concepts and propositions with a *network structure*. In fact, neither of these representational assumptions are very constraining, but the availability of an agreed upon unit of text analysis (i.e., propositions) and a way to conceptualize how those units might be represented in a coherent structure (i.e., a network) were important theoretical developments.

Second, Haviland and Clark (1974) published an elegant paper that made an important theoretical contribution and served as a model for investigating on-line comprehension processes. Haviland and Clark were interested in a hypothesis about how readers integrate related statements in a text. They proposed that writers use syntactic devices to distinguish "new" information in a sentence from previously established, or "given," information. They further hypothesized that readers are sensitive to this distinction and follow a general strategy of locating the referent of the given information in memory, then "connecting" the new information with that context. In this manner, readers systematically construct an integrated representation of a text. As we will elaborate later, this simple model provided a foundation for subsequent theorizing about the nature of comprehension processing and its memory products.

The third important development in the 1970s was the introduction of computer-controlled eye-tracking procedures to study reading processes (McConkie & Rayner, 1975; Rayner, 1975). Perhaps the most important consequence of this development was the interest in text processing that it sparked. However, the methodological advantages of eye-tracking procedures have also been critical in advancing research on text processing. Rayner and McConkie studied readers' perceptual spans using eye-tracking equipment with extremely precise spatial and temporal sensitivity. They interfaced the eye-tracker with a computer so that text display could be made contingent on the eye movements of the reader. For instance, the system could detect the character within a word that a reader was fixating at a given point in time, then make a change in the visual display between the time the reader initiated a forward saccade and the time that the eye movement landed at its destination. With this methodology, Rayner and McConkie (1976) were able to determine in detail the characteristics of readers' perceptual spans (e.g., location, width, what types of information are processed in the visual periphery). Although virtually all of the early research using this eye-tracking technology focused on the nature of the initial stages of processing during reading (e.g., basic visual processes and word encoding), eye-tracking procedures have been increasingly used to study processes more closely associated with comprehension, such as sentence parsing (Frazier & Rayner, 1982; Rayner & Frazier, 1987) and inferential processing (O'Brien, Shank, Myers, & Rayner, 1988; Garrod, O'Brien, Morris, & Rayner, 1990).

The final key development in the 1970s was the publication of an ambituous theory of text processing by Kintsch and van Dijk (1978). Their theory integrated ideas about knowledge representation borrowed from the computer-inspired models of memory (e.g., Anderson, 1976; Anderson & Bower, 1973; Kintsch, 1974) with ideas about the nature of on-line comprehension processes (e.g., Haviland & Clark, 1974) and the basic cognitive constraints within which those processes operate (e.g., working memory). It simultaneously managed to tackle a broad range of issues concerning reading, while generating concrete, testable hypotheses about key mechanisms and processes underlying reading. The paper also forms a curious bridge between prior and subsequent research on text processing. Despite the theory's clear statements about the nature of processing occurring during reading, Kintsch and van Dijk's own empirical tests of the model followed the pattern of earlier research in that their tests relied on memory data. It was largely left to subsequent researchers to test the implications of the theory for on-line processing (e.g., Fletcher, 1981, 1986).

The on-line study of comprehension processes

In 1980, Just and Carpenter (1980) presented a theory of text processing that shared the general theoretical framework introduced by Kintsch and van Dijk (1978), but focused directly on the study of on-line reading processes. Just and Carpenter were particularly interested in using eye-tracking methodologies to study reading and they presented a theoretical analysis of how eye movements reflect the cognitive processes of reading. They proposed the "immediacy principle," which states that a reader does not make an eye movement until completion of all the processing that can be accomplished during the current fixation. This hypothesis of a tight relation between cognitive processing and eye movements allowed investigators to closely examine the lexical, syntactic, and semantic processes operating *during* text comprehension. It encouraged theorists to develop models of processing that pinpoint the event within a text (e.g., word, clause boundary) that should trigger a specific process and it encouraged researchers to develop procedures to test the detailed models. The fruits of their efforts included improved methods for studying on-line text processing and a coherent body of empirical findings concerning the nature of on-line comprehension processes.

Methodological developments

The work of Just and Carpenter exemplifies what is, in many respects, the most important contribution of cognitive science to the study of reading comprehension, namely, an emphasis on studying the *processes* of comprehension *as they occur*. To study comprehension processing on-line, cognitive scientists have relied primarily on two categories of procedures.

One category of procedures allows relatively uninterrupted reading of the text and measures the time it takes to process particular target items in a text. A simple and frequently used procedure has been to have readers do self-paced reading of a text that is presented a sentence at a time while the time to read each sentence is recorded (Cirilo & Foss, 1980; Haviland & Clark, 1974). Variations in sentence reading time are interpreted as reflecting the relative difficulty of comprehending the sentence (i.e., relating it to prior context) when suitable controls are employed to rule out other interpretations. The most sophisticated procedure in this category uses computer-controlled eye-tracking techniques (Rayner, 1975). These allow extremely detailed measurement of reading behavior, including fixation locations and durations at the level of individual words, the direction of eye movements (i.e., forward saccade vs. regressive eye movement), the origin or destination of particular eye movements, and so on.

The sentence-reading and eye-tracking procedures have been particularly useful in tracking the time course of processing during reading, but they generally do not reveal what specific information is available to, or being processed by, the reader at a particular point in time. For instance, slow processing of a statement may indicate difficulty in determining the relation of the statement to prior text information, but it does not reveal what information is being activated by the memory search to locate an appropriate context for integrating the statement. For this purpose, researchers often use a second category of procedures, probe techniques. In a probe task, reading is interrupted at a critical point by the presentation of a word, phrase, or sentence to which a response is required. Among the more typical procedures, a word is presented and the reader must either name the word aloud (e.g., O'Brien, Plewes, & Albrecht, 1990), make a lexical decision (e.g., McKoon, Ratcliff, & Ward, 1994), or determine whether the word occurred earlier in the text (e.g., McKoon, Gerrig, & Greene, 1996). The logic of the procedure is based on studies of priming in memory and attention tasks. If the concept designated by the probe word is highly available or "active" as a result of processing occurring during reading, then the time to make a response to the probe word will be relatively short; if the probed concept is not active, response time to the probe word will be relatively long. Thus, probe procedures have provided a means of tracking the availability of specific information during reading.

Empirical findings

The application of the various methods for the on-line study of comprehension processes has resulted in a great number of empirical findings; we will summarize these findings only in the most general terms.

Research in cognitive science has concentrated heavily on the role and nature of inferential processes during reading. This emphasis follows from

the dominant theoretical framework, in which text comprehension is defined as the process of constructing a connected memory representation (Just & Carpenter, 1980; Kintsch, 1988; Kintsch & van Dijk, 1978; McKoon & Ratcliff, 1992; van den Broek, 1990). In this view, inferences play a central role in constructing the connections among the concepts and propositions in the memory representation. Most of the research on inferences has focused on the questions: What types of inferences do readers routinely make? And under what circumstances do readers make those inferences?

One type of inference that has received considerable attention concerns those that establish referential relations during reading. Anaphoric devices, such as pronominal reference and ellipsis, have been studied extensively. It is well established that readers are conscientious in resolving references as soon as they are introduced in a text. This conclusion is indicated by findings that the time it takes readers to process an anaphor (Ehrlich & Rayner, 1983) or a sentence containing an anaphor (Sanford, Garrod, & Boyle, 1977) depends on factors affecting the ease with which the referent of the anaphor can be unambiguously identified. The conclusion is also supported by results from studies in which probe procedures are used to track the availability of a specific target concept during reading. For instance, a concept that is not active at some point during reading will become highly avaiiable as a result of processing of an anaphoric reference to the concept (O'Brien & Myers, 1987).

A second set of inferences that has been investigated extensively concerns causal inferences during the reading of narratives. Trabasso and van den Broek (Trabasso, Secco, & van den Broek, 1984; Trabasso & Suh, 1993; van den Broek, 1990) proposed that narratives are organized around a sequence of causally related events that are motivated by the goals of the characters. Their theory provides a means for analyzing narratives with respect to the causal connections that exist among pairs of events in a story. The result is a network of causally related events that has been demonstrated to represent an important source of both local and global coherence in narrative. According to Trabasso and van den Broek's analyses, events in a narrative vary with respect to (a) the number of causal connections they have to other events and (b) whether they lie on the "causal chain" that leads from the initial goal of the protagonist to the final outcome of the story. Both the causal chain status of an event and the number of causal connections of an event have been found to be important determinants of the probability of recall of the event (Goldman & Varnhagen, 1986; Trabasso & van den Broek, 1985; van den Broek, Lorch, & Thurlow, in press) and of readers' ratings of the importance of the event (Trabasso & Sperry, 1985; van den Broek, 1988). The causal structure of narrative is also related to on-line measures of text processing. For instance, the time to comprehend a statement is a function of the ease of identifying a causally sufficient prior event in the story (Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Fletcher

& Bloom, 1988). In addition, the time to identify the antecedent of an anaphoric reference in a narrative is a function of the distance of the reference from the antecedent as defined by a causal analysis of the narrative (O'Brien & Myers, 1987). In summary, much empirical evidence indicates that readers systematically identify and represent causal relations among story events as a central component of the process of comprehending a narrative. Further, at subsequent recall, the retrieval of information from readers' text representations is strongly influenced by the causal structure of the text representation.

Although referential and causal inferences have received the lion's share of research attention, other types of inferences have been studied as well. These include spatial inferences (Zwaan & van Oostendorp, 1993), instrumental inferences (Singer, 1980) predictive or elaborative inferences (Singer & Ferreira, 1983; Potts, Keenan & Golding, 1988; McKoon & Ratcliff, 1986; Klin & Myers, 1993), and others (for a review, see Graesser, Bertus, & Magliano, 1995). In general, these types of inferences appear to be made less frequently than referential and causal inferences, but the conditions under which they are made appear to be regular. Three general, related principles capture much of what we know about inferencing during reading. First, readers monitor text content with respect to its internal consistency (Kamas & Reder, 1995; O'Brien, 1995; van den Broek, Risden, & Husebye-Hartmann, 1995). They are sensitive both to contradictions and to "gaps" in their mental representations. If readers encounter a statement that contradicts information established earlier in a text, they are generally slow to process the statement and they attempt to construct some resolution of the contradiction (Albrecht & O'Brien, 1993; O'Brien & Albrecht, 1992; Myers, O'Brien, Albrecht, & Mason, 1994). If they encounter a statement that cannot be readily related to the immediate context in the text, they search their memories for a suitable prior context or construct an inference that bridges the gap (Bloom *et al*., 1990; van den Broek, Rohleder, & Navarez, 1996).

Second, readers are much more likely to make inferences when the text requires them to do so (i.e., when coherence breaks or contradictions are encountered) than when comprehension of the text does not mandate an inference. For instance, referential inferences are probably common because comprehension of a sentence requires resolution of all anaphoric references. In contrast, understanding that the victim in a murder mystery was stabbed to death does not require the inference that the murderer used a knife, so that inference is generally not made (Dosher & Corbett, 1982; Keenan & Jennings, 1995). However, if a reference is subsequently made to "the bloody knife," the reader is then required to infer that the murder weapon was a knife (O'Brien *et al*., 1988).

Finally, readers are also likely to make an inference if the appropriate inference is highly constrained by the context. For example, if the text contains a passage describing an animal with many of the characteristics of a skunk but does not explicitly mention a skunk, the reader is likely to infer

that the animal is a skunk even though the inference is unnecessary (O'Brien & Albrecht, 1991; van den Broek, 1990; Vonk & Noordman, 1990).

In sum, substantial progress has been made toward distinguishing various types of inferences, identifying the conditions under which they are made, and describing the mechanisms underlying them.

Analyzing the contributions of cognitive science

Having reviewed the development of cognitive research on text comprehension, we will now elaborate on what we consider the most important contributions of cognitive science to date. These are: (1) the development of theories of representation of knowledge; (2) the view of comprehension as the construction of a memory representation; and (3) the emphasis on detailed analyses of comprehension processes and the development of methodologies to study those processes on-line. This discussion will serve to illuminate the reasons for cognitive science's successes in describing the reading process and will set the stage for considering the directions future research should take.

The significance of the development of theories of representation

The development of theories of knowledge representation in the early 1970s was critical to the subsequent development of cognitive research on text comprehension in several respects: (1) The theories provided a basic unit of analysis of text content in the form of the proposition; (2) they provided a metaphor for visualizing relations among meaning units in the form of network structures; and (3) they provided the foundation for theorizing about process.

The proposition as the basic unit of meaning

A reader who is asked to recall a text of any length will accurately recall the meaning of some parts of the text, will distort the meaning of other parts, and will omit much of the text content from recall. To begin to understand the transformations that occur from reading to recall, researchers required a principled means of analyzing text content and the content of a reader's recall. However, prior to the 1970s no such tool was available. In lieu of a theoretical basis for identifying the basic meaning components of a text, researchers simply required verbatim recall and counted how many words or sentences in a text were accurately recalled, or they identified "idea units" in an *ad hoc* manner, or they opted for an empirical means of identifying meaning units in a text (Johnson, 1972). The introduction of the proposition as the hypothetical basic unit of meaning in a text was an important theoretical and methodological advance in text comprehension research (Frederiksen, 1975; Kintsch, 1974; Meyer, 1975).

Propositional analysis has a long history in logic and the advantages of the proposition as a unit of meaning are well known (see Anderson & Bower, 1973). The fact that propositions are relatively well-defined units of analysis has the important methodological consequence that scoring systems based on propositional analyses have generally proven to be quite reliable. In addition, the availability of propositional scoring systems provided the foundation for replicability of empirical studies because they established a common reference point for theoretical analyses of text content. Finally, although propositions are cumbersome with respect to representing some nuances of meaning, they are a powerful and flexible representational formalism that has generally sufficed for the purposes of most researchers. At the same time, the adoption of the proposition as a basic unit of analysis does not appear to be a very constricting theoretical commitment in the sense that it is unlikely that research based on propositional analyses would be undermined should the proposition be demonstrated to be an "incorrect" representational assumption.

Networks as metaphors for text structure

Texts and memory representations are not simple lists of propositions; they are organized structures. The theories of knowledge representation developed in the 1970s devoted a great deal of attention to the question of how comprehenders organize propositions in their memory representations. Frederiksen's (1975) theory detailed multiple relations that may be represented among propositions, including spatial, temporal, logical, and causal relations. Schank (1975) was particularly interested in knowledge structures organized by causal relations and knowledge structures organized by temporal and pragmatic relations (i.e., scripts). Kintsch and van Dijk (1978) and Anderson and Bower (1973) emphasized the role of coreferential relations in guiding readers' constructions of text representations. Meyer (1975) was most interested in the nature of the rhetorical relations that organize expository text at the most superordinate levels of the text representation. Despite their different goals, all of these theorists chose to represent relations among propositions as network structures.

As the network metaphor has been applied by cognitive scientists, it embodies the assumption that readers connect propositions into their memory representations as they read. It also provides an efficient means for representing hierarchical relations among propositions, although it not restricted to representing hierarchical relations. The network metaphor does not, in itself, have strong theoretical implications, however, because it does not specify the nature of the connections readers construct. The main constraints on the final form of a representation derive from hypotheses provided by the theorist concerning (1) the types of relations that will be represented by a reader and (2) processing constraints operating during reading, such as

working memory capacity (Kintsch & van Dijk, 1978). Perhaps because the assumption of a network structure entails few theoretical commitments, it has proven a productive framework for theorizing about the processes underlying the construction of a memory representation by the reader. On the one hand, it encourages the theorist to visualize comprehension as the construction of a network of related propositions. On the other hand, it requires the theorist to specify the principles and constraints that guide the construction process. Certainly, the metaphor continues to be a useful one for cognitive theorists (e.g., Kintsch, 1988; van den Broek, 1990).

A foundation for theorizing about processes

A theory of representation may serve as a theory of the *products* of comprehension, but it is not a theory of the *processes* of comprehending a text. There are many alternative processing models that might be proposed for constructing a propositional network representation of a text. Thus, hypothesizing that readers construct such a representation does not tell us much about how the representation is constructed. However, the hypothesis does have some important implications for the nature of such processes. First, it implies that an early stage in processing involves parsing the sentences of a text into their component propositions. Second, it specifies the nature of the "building blocks" for the mental representation, which enables theorists to become much more concrete in modeling how readers process a text. Third, it implies that an important subset of comprehension processes are those that are involved in identifying and representing relations among the propositions (i.e., "connecting up" the network). Thus, it suggests a metaphor for the process of text comprehension, namely, that the process of comprehending a text may be viewed as a process of constructing a memory representation.

The significance of viewing comprehension as memory construction

Most of the early propositional theories were concerned primarily with issues of the form of the representation of knowledge and how those representations are accessed and retrieved from memory. Haviland and Clark's (1974) seminal paper was concerned with issues of how readers construct knowledge representations in the first place. They proposed that readers interpret various linguistic expressions as instructions about how to integrate new information in a text with relevant prior information from the text. Their given-new strategy is a simple model of the nature of the memory operations that are performed in the process of interpreting a statement and constructing an appropriate memory representation. The perspective that the process of comprehending may be viewed as the construction of a memory representation has important implications for cognitive science research on text comprehension.

A definition of comprehension

In an important sense, viewing reading comprehension as memory construction provided researchers with a definition of "comprehension." Previously, researchers often implicitly defined comprehension as whatever their assessment instrument measured (e.g., number of ideas recalled in a free recall task; number correct on a multiple-choice, recognition test of memory for text content) (see Carroll, 1972, for a review and critique). Viewing reading as the construction of a memory representation defines comprehension in terms of the coherence of the representation the reader constructs and —depending upon the reader's goal—the relation between the reader's representation and the representation intended by the author: A reader comprehends a text to the extent that the reader's representation captures the local and global coherence relations intended by the author.

The perspective of reading comprehension as memory construction has psychological validity in the sense that it appears to correspond to how readers define comprehension for themselves. That is, readers are satisfied that they have understood a text to the extent that their representation of the text is coherent. Further, readers evaluate their ongoing comprehension attempts—as well as the end-product of their reading—with respect to the coherence of the memory representation they are assembling. There is substantial empirical evidence (see McKoon & Ratcliff, 1992; Graesser, Singer, & Trabasso, 1994; Singer, Graesser, & Trabasso, 1994; Gernsbacher, 1990; van den Broek *et al*., 1995) that readers' criterion for whether they have adequately understood a statement is whether the information in the statement can be "connected into" the representation they have constructed to that point during reading. In summary, viewing comprehension as memory construction has proven useful for the theorist in large part because it defines "comprehension" and because it corresponds to the readers' perspective of their task.

A broad theoretical perspective

Viewing text comprehension as memory construction places the topic of reading comprehension in a broad theoretical framework and, thus, brings to bear on the study of reading the full arsenal of concepts in cognitive science. In particular, the fruits of research in the domain of memory have greatly influenced cognitive theories of reading comprehension.

One critical contribution to theorizing about reading comprehension is the observation that readers' attempts to understand text statements as they read are constantly constrained by the limits of working memory (Kintsch & van Dijk, 1978; Just & Carpenter, 1992). Individual differences in working memory capacity have been demonstrated to be a good predictor of variation in both overall reading ability (Daneman & Carpenter, 1980) and specific

reading skills (Daneman & Carpenter, 1983). In addition, the comprehensibility of a text can be well predicted by an analysis of the demands it makes of readers' working memories (Britton & Gulgoz, 1991; Miller & Kintsch, 1980).

In large part because of the constraints working memory places on readers' comprehension efforts, it is critical that readers be able to efficiently access information in long-term memory. Thus, models of memory search and retrieval constitute a second important contribution of memory research to theorizing about reading comprehension (Gillund & Shiffrin, 1984; Ratcliff, 1978). Modeling comprehension processes in terms of a memory search of a text representation has produced very successful accounts of basic empirical findings in reading comprehension research, including the resolution of anaphors (Kintsch, 1988; McKoon & Ratcliff, 1992; O'Brien, 1995) and the establishment of causal coherence (van den Broek, Risden, Fletcher, & Thurlow, 1996).

Although the emphasis in this paper is on the impact that cognitive science has had on our understanding of reading, it should be noted that the field of reading research, in turn, influences the broader field of cognitive science. Our understanding of many basic cognitive processes is based largely on empirical procedures that attempt to isolate the process of interest as much as possible. Reading is a complex behavior that requires the smooth integration of many basic cognitive structures and processes, including attention, working memory, long-term memory, and various linguistic processes. Thus, the study of reading affords cognitive science an opportunity to test the sufficiency of its basic theoretical constructs with respect to accounts of complex behavior. One clear lesson that may be derived from cognitive research on reading is that readers have rapid access to a great deal of information at any given point in time during reading (Albrecht & O'Brien, 1993; Myers *et al.*, 1994; O'Brien & Albrecht, 1992). This observation is already having an influence on basic theories of memory, particularly on the concept of working memory (Ericsson & Kintsch, 1995).

Linguistic devices as processing instructions

A corollary of viewing comprehension as memory construction is the notion that the text, itself, constitutes a set of "instructions" that trigger and direct the cognitive operations that result in comprehension. Haviland and Clark's (1974) most important contribution was, perhaps, the basic insight that various linguistic devices are interpreted by readers as instructions about how to integrate a text statement into their memory representations. They showed that the information necessary to determine how a specific sentence is related to prior context is often conveyed by a single word (e.g., a pronoun) or the sentence's grammatical construction (i.e., marking "given" vs. "new" information). Further, they demonstrated that readers use such

information on-line to direct comprehension processes. Specifically, Haviland and Clark showed that the time required to comprehend a sentence is predictable from (1) careful identification of what information is required to "integrate" two statements, in conjunction with (2) an analysis of the availability of that information in memory. These insights complemented the existing, predominantly propositional, theories of knowledge representation. These theories did not address questions concerning early stages of sentence processing (e.g., word identification) and their potential influence on the processes by which relations are identified between sentences. Rather, they assumed a parsing mechanism that analyzed sentences into their component propositions and they picked up their analyses at that point, identifying relations among propositions based on shared arguments (Anderson & Bower, 1973; Kintsch, 1974; Kintsch & van Dijk, 1978). Thus, Haviland and Clark provided the basis for accounts of reading comprehension in which word and sentence processing are integrated.

The significance of detailed models of processing

Prior to the early 1970s, theorists characterized the nature of reading processes only in the most general terms. Typical controversies concerned the extent of top-down vs. bottom-up processing during reading and whether inferences revealed in memory tests were made during reading or at the time of test (Spiro, 1977; Royer, 1977). Little attempt was made to separate different processes or to distinguish different types of inferences. This situation changed in the mid-1970s with the development of detailed processing theories and methodologies to test them. Theoretical and methodological advances have proceeded hand in hand. The development of the eye-tracking paradigm (Rayner, 1975) led to closer consideration of reading processes because the fine grain of the data encouraged detailed analyses. In turn, the availability of theories that made testable predictions about on-line processing (e.g., Haviland & Clark, 1974; Kintsch & van Dijk, 1978) led researchers to adapt procedures from other domains to test the theoretical predictions (e.g., Fletcher & Bloom, 1988; Fletcher, 1981, 1986).

A preference for detailed theories of processing is a hallmark of the cognitive science approach and, to a great extent, it is the source of the theoretical and empirical contributions of cognitive science to our current understanding of text comprehension. In particular, much of the progress that has been made in understanding inferential processes is attributable to the development and testing of detailed models. We already have summarized the basic findings from that literature, so here we focus on the basic research strategy that characterizes much cognitive science research on text processing.

Cognitive scientists have generally taken an experimental approach to the study of reading processes. First, theorists have identified important

functions comprehenders must execute during reading. For example, readers must resolve references to established entities in a text and they must make various types of inferences to fill in "gaps" in the text. Next, for each function identified as a necessary component of comprehension processing, theorists have tried to specify text conditions that should invoke that function. For instance, various devices may be used by an author to establish referential connections between entities in a text, including the use of pronouns, definite reference, and ellipsis. Once text conditions of potential interest are identified, the researcher attempts to isolate the conditions of interest, generally by writing carefully controlled texts. Theories are developed to address the nature of processing under those conditions and to suggest variables that may control processing. Finally, procedures are developed to test the theories of processing under laboratory conditions.

The attention to detail that characterizes cognitive science research is necessary for progress toward understanding text comprehension. Writers do not express themselves in arbitrary ways. They try to induce the members of their audience to construct a representation that is faithful to the message they wish to convey. Sentences are structured to communicate a particular perspective; many different devices are used to communicate referential relations within the text; authors assume that their readers possess certain background knowledge. Writers construct their texts with close attention to the details of their writing on the assumption that readers will attend and respond appropriately to those details. It behooves the theorist to strive for a similarly detailed analysis of the reader's task. Each reference in a text must be resolved by the reader, so an understanding of reading processes must address how the reader determines the reference of a pronoun, or a definite reference, or an elliptical reference. Grammatical structures vary in the ways in which they distribute emphasis across sentence content, so the theorist must consider the effect on the reader's processing of the text. In short, a sufficient account of comprehension processing must address the complexity of the reader's task in all its detail. By providing detailed, rather than global, superficial accounts of the reading process, cognitive science allows us to address issues which previously were deemed beyond study.

Directions for future research

Although cognitive science has had an important influence on text comprehension research in the past 2 decades, we believe that its most important contributions are yet to come. In this section, we discuss the directions in which we hope research will proceed. We first indicate areas in which research is already being conducted and in which major advances are likely to occur. We then identify issues that currently are not being investigated, but that merit attention in future research.

Current and developing trends

Cognitive research on reading comprehension continues to evolve in several directions: (1) Current theoretical issues of debate will shape future research in cognitive science. (2) Connectionist modeling has been introduced into cognitive theories of reading comprehension (Just & Carpenter, 1992; Kintsch, 1988) and is likely to become an increasingly important tool for theorists. (3) There are several topics that are receiving increasing attention from researchers.

Theoretical issues

Several issues in reading research are currently debated, sometimes hotly. Two of these issues have particularly sparked research. The first concerns the nature and extent of inferencing in which readers routinely engage and the corresponding richness of their text representations. We have seen that there is a great deal of empirical evidence demonstrating that readers are conscientious in constructing a text representation that maintains local coherence. If a statement is encountered that cannot be related to the immediate context in which it occurs, readers search their long-term memories for an appropriate interpretive context and/or generate an inference that relates the problematic statement to an earlier context. Theorists agree that readers routinely make the inferences necessary to maintain a locally coherent mental representation of a text. However, they disagree with respect to the extent of additional inferential processing that is routinely done by readers.

At one extreme, McKoon and Ratcliff (1992) have proposed that readers typically make *only* those inferences necessary to maintain a locally coherent text representation and those inferences based on "highly available" information. At the other extreme, Graesser, Singer, and Trabasso (1994; Singer *et al*., 1994) have proposed that readers also routinely make various inferences necessary to construct a globally coherent text representation. Although the "minimalist" versus "constructionist" debate has often been framed in terms of the question of whether readers make "global inferences," the term "global" turns out to be quite ambiguous. It has been used to refer both to surface distance in a text and to processing of what Kintsch and van Dijk (1978) termed the "macrostructure" of a text. In fact, there is little debate that "horizontal" inferences (i.e., one-to-one connections between two concepts or sentences) often span relatively long surface distances in a text (McKoon & Ratcliff, 1992; Myers *et al*., 1994; O'Brien & Albrecht, 1992). Rather, the controversy concerns whether readers routinely make "vertical" inferences; that is, whether they process hierarchical relations involving many-to-one connections among concepts or statements in a text.

The debate over whether readers routinely construct vertical inferences may ultimately reduce to a debate over what constitutes "routine" reading

situations. At a fundamental level, the minimalist/constructionist debate contrasts different perspectives on reading. The minimalists view comprehension processing as involving a core set of nonstrategic processes that are unaffected by readers' goals and motivations or the type of materials being read. They do not deny that many strategic processes are involved in reading, but they are not interested in these processes and dismiss them as specific to "special" goals of readers. Constructionists view reading as fundamentally goal-directed and highly strategic in nature. In fact, the two camps define different domains of interest and may ultimately prove to be compatible theoretical positions (van den Broek, Fletcher, & Risden, 1993; van den Broek *et al.*, 1995). At present, however, the points of contrast that have been drawn between them focus on critical questions concerning reading processes.

The second issue of interest concerns a distinction between two levels of text representation; a textbase representation vs. a situation model, or mental model representation. It is claimed that readers form both types of representations, although the representations compete for processing resources so that there may be tradeoffs in the degree to which each representation is developed (Schmalhofer & Glavanov, 1986; van Dijk & Kintsch, 1983).

As the distinction was originally proposed, a textbase is a representation of a text, whereas a situation model is a representation of what a text is about. Kintsch and van Dijk's (1978) theory was concerned primarily with the nature of a reader's textbase representation and how the representation was constructed during reading (i.e., the leading edge strategy). The theory analyzed relations among propositions in terms of their argument overlap and emphasized the roles of hierarchical relations and recency in the construction of a memory representation. In short, a textbase representation captures the linguistic relations among propositions in a text. In contrast, a situation model focuses on a reader's attempts to connect statements in a text based on their relations in a possible world. For example, whereas a textbase representation of the description of a floor plan of a house would be organized according to the sequence of descriptions of rooms, a situation model representation would be organized according to the spatial relations between the rooms (i.e., would correspond more closely to the actual floor plan than to a verbal description of the plan), (Morrow, Greenspan, & Bower, 1987). As another example, events in a narrative are understood in the context of the personalities and motivations of the story's characters (O'Brien *et al.*, 1990; Trabasso *et al.*, 1984).

Currently, the distinction between a textbase vs. situation model representation is a matter of debate. Both constructs are complex and not sufficiently specified. However, research motivated by the distinction has led in at least two fruitful directions. One consequence of the distinction has been increased attention to the role of a reader's background knowledge during text processing. For example, how is background knowledge retrieved and used to direct the comprehension process? And how is new information

acquired from a text integrated with knowledge already stored in memory? This shift in perspective is in sharp contrast to Kintsch and van Dijk's (1978) original model, which explicitly excluded from consideration the role of background knowledge.

The other consequence of the distinction between textbase and situation model representations is closely related to the first. Current research is increasingly concerned with the issue of how readers identify relations among text statements based on the meaning and implications of the statements. For example, how does a reader quickly detect a contradiction between an event that is currently being processed and information presented much earlier in the text (Myers *et al*., 1994; O'Brien & Albrecht, 1992)? Again, this is in contrast to the Kintsch and van Dijk's emphasis on the role of simple argument overlap in identifying relations among propositions within a text.

In sum, both the minimalist/constructionist debate and the textbase/situation model distinction are unresolved controversies. However, both controversies are motivating researchers to examine some central questions concerning reading comprehension.

Connectionist models of reading processes

An important development in cognitive theorizing about reading comprehension has been the formalization of two ambitious theories in the form of connectionist models (Kintsch, 1988; Just & Carpenter, 1992). Connectionist models have proven quite useful for modeling in other domains that involve problem-solving processes operating under multiple constraints (Bechtel & Abrahamsen, 1991; Holyoak, 1991). This is certainly an apt description of the process of constructing a text representation. With each new sentence in a text, readers must retrieve from memory both text information and prior knowledge that may be relevant to the interpretation of the sentence. They must then determine what activated information is most relevant and how to integrate the new information with their mental representation of the text. The underlying processes must operate rapidly and continuously while under the constraints imposed by working memory. Connectionist modeling is currently the most powerful tool available for theorizing about such complex, interactive processing (Holyoak, 1991).

The conventional research strategy in cognitive science, in general, and reading research, in particular, has been to devise tasks designed to isolate specific cognitive processes for study under controlled, experimental conditions. Although this reductionist strategy is tried and true, it must be balanced by consideration of how the component processes might be reintegrated into a broad theory of reading comprehension processes. Computational modeling offers one means of addressing the reintegration problem. As computer simulations of reading, computational models require the type of detailed theoretical analyses of processes that cognitive scientists prefer. At

the same time, they require the theorist to attend to issues of sufficiency. In particular, the theorist must model the interaction of multiple processes, as well as the individual processes themselves.

A comprehensive theory of reading comprehension must simultaneously explain the wide range of behaviors associated with reading: It must be able to account for a variety of on-line processes (e.g., the resolution of anaphoric or causal ambiguities; the activation of background knowledge), as well as the effects of such processing on the products of reading (e.g., recall of text content; question-answering). Consequently, a sufficient theory of on-line comprehension processing will be very complex. The predictions of models of component comprehension processes (e.g., how pronominal references are resolved on-line) may be straightforward when considered in isolation, but it may not be possible to deduce the behavior of the overall model from a consideration of its individual components. Computational modeling provides a framework for implementing many processes simultaneously and observing the overall behavior of the model, as well as the behavior of the individual components in the context of the complete model. Thus, connectionist models provide a means of testing the sufficiency of a theory of reading.

Related to the preceding point, computational models can be used to examine the nature of interactions among hypothetical processes. First, in the course of developing a computational model, the theorist is forced to consider precisely how the model's component processes will work together. This exercise encourages the development of alternative models of the interaction of component processes. Second, once a particular model is developed, running the simulation may lead to insights into the nature of interactions among processes. The behavior of the model may not conform to the expectations of the theorist because the nature of interactions among components of the model may not be fully anticipated. Analyzing the source of unexpected behavior by the model may lead to insights into reading behavior. Of course, a researcher need not depend on unpredicted interactions; a running simulation may also be used to systematically explore alternative models of interactions among components.

Finally, computational models typically possess general parameters that have psychological interpretations. In Construction-Integration theory (Kintsch, 1988), for example, there is a parameter corresponding to the extent of search of long term memory before integration occurs. Psychologically, the search parameter could be affected by a reader's goals and motivation and is likely to influence both the elaboration and coherence of the mental model the reader constructs and the degree to which the mental model is integrated with background knowledge. The availability of a computational model permits the researcher to explore the implications of different parameter settings.

In sum, connectionist modeling will be an increasingly important tool in reading research. However, as its application to reading is continued

and expanded, it is important that modelers keep firmly in sight the goal of understanding reading comprehension. Model sufficiency should not be achieved at the cost of transparency. If the reasons for a model's behavior cannot be identified unambiguously, then the model cannot very well enlighten the (possible) reasons for corresponding human behavior. Processes implemented in a model should represent hypotheses about corresponding psychological processes and parameters of a model should have psychological interpretations. A model should produce testable predictions about reading behavior and the predictions should be tested against human performance.

Developing topics of research

As in any active domain of research, new topics and issues are constantly being identified and pursued by investigators. In this section, we note two significant current trends.

In the 1970s, researchers were very much concerned with questions about the nature of the representation that resulted from reading a text. In the 1980s, attention shifted to questions about the nature of the on-line processes that construct a mental representation during reading. More recently, investigators have attempted to relate on-line processes to the resulting mental representation. At an empirical level, researchers have been increasingly concerned with demonstrating that inferences detected by on-line processing measures have corresponding effects on readers' text representations, as indicated by off-line memory measures (e.g., Graesser & Clark, 1975; Klin, 1995; O'Brien & Myers, 1987; Trabasso & Suh, 1993; van den Broek, 1990). More importantly, recent theoretical treatments explicitly address the relationships between on-line processing and subsequent, off-line access to the text representation (Goldman, Varma, & Cote, 1995; Kintsch, 1988; St. John, 1992; van den Brock *et al*., 1996). These theories describe the dynamic processes by which on-line processes interact with subsequent off-line processes. Their goal is discovery of the mechanisms by which on-line and off-line processing is connected.

We encourage continued attention to the relation between process and product in reading. A reader processes a text, in large part, to construct a representation—a representation that may then be used to support other purposes. Thus, cognitive theories must specify not only what processing is done during reading, but the representational function of these processes, as well. Doing so is not only essential to constructing a complete account of reading, it also has important potential practical implications: Once we understand how what a reader does *during* reading affects what he or she can access *after* reading, we can better design instructional approaches to teach effective comprehension strategies and remedial approaches to identify and correct ineffective strategies.

The second research trend to be discussed is the study of the affective responses of readers as they read. Traditionally, cognitive science has paid little attention to this topic. However, several investigators have recently begun to address this omission (Gernsbacher, Goldsmith, & Robertson, 1992; Stein & Levine, 1990; Stein, Liwag, & Wade, 1995; Trabasso & Magliano, 1995). There are two general reasons that it is important to study the affective responses of readers. First, for some text genres, involving the reader affectively is a primary goal of the author and an important motivation for the reader (Brewer, 1980; Lorch, Klusewitz, & Lorch, 1995; Lorch, Lorch, & Klusewitz, 1993). In the case of narrative, for example, the genre is defined, in part, by the inclusion of surprising events. Thus, a complete understanding of the processes of text comprehension must include an understanding of the reader's affective responses. Second, for all text genres, the motivation and interest of the reader surely are important factors influencing the extent of cognitive processing of a text (Hidi, 1990; Shirey & Anderson, 1988). This is the case even for expository texts. Although authors of exposition are usually more concerned with communicating information than with involving readers affectively, the involvement of the reader is probably a critical determinant of the effectiveness of the communication. The extent of affective involvement of the reader will influence the nature and extent of cognitive processing of a text, which, in turn, will determine what is learned from the text.

Topics meriting more systematic investigation

Cognitive science has already had impressive successes and the trends in current research promise to yield even more insights. Yet, there are important topics and issues that have not received adequate attention. We address some of those topics in this section.

The representation and processing of expository texts

There has been insufficient attention in the cognitive literature to processing of nonnarrative text. We hasten to acknowledge that there are many reports in cognitive journals of research on the comprehension of technical and seientific texts and other forms of exposition (Dee-Lucas & Larkin, 1988; Kieras, 1981; Royer, Carlo, Dufresne, & Mestre, 1996). However, the research on narrative comprehension is more extensive and *more programmatic* as a field than research on other text genres. As a result, current cognitive theories of text comprehension are based primarily on the empirical literature concerning inferential processing of narrative. This state of affairs raises the question whether findings from studies on comprehension of narratives generalize to other types of texts. It is likely that some findings do generalize whereas others do not. For example, processing of referential relations may

not depend on the genre. On the other hand, its tight causal structure may result in narrative being processed in ways that are systematically different from those that occur during reading of expositions (cf. McDaniel, Einstein, Dunay, & Cobb, 1986). Thus, from a theoretical perspective, it is important to study the comprehension of expository text. It is also important from a pragmatic perspective because various types of expository text are very common in educational and work settings.

What might be the most fruitful direction in which to pursue programmatic research on exposition? The major reason for the success in studying processing of narrative is that we have a relatively good theoretical understanding of the structure and content of narrative. Theories of narrative representation (e.g., Trabasso *et al*., 1984) have, in turn, provided a foundation on which to build theories of processing (van den Broek, 1990). No comparably well-developed theories of the representation of exposition exist. This is surely due, in part, to the fact that "exposition" is not nearly as homogeneous in structure as narrative. Thus, perhaps the first step in creating a theory of representation of exposition should be to distinguish subtypes, each of which has its own characteristic structure (e.g., newspaper articles, scientific journal article, instructions, description). Along these lines, Brewer (1980) offered a preliminary classification of discourse types based on considerations of the underlying psychological representation of different text genres and the purposes for which authors write. Assuming that expository subtypes can be identified at a level of categorization similar to narrative (i.e., with a similarly stereotypical structure), a representational theory must then be formulated for each subtype of interest. Taking a lesson from the history of theories of narrative, the type of representational theory that is likely to prove most useful is one that provides an analysis of how to construct an appropriate mental model from the text, as opposed to simply analyzing the "grammatical" structure of the text. The availability of a well-developed representational theory for a specific expository subtype would serve as the starting point for the systematic development and testing of processing theories, as it did with research on narrative. In addition, the availability of representational theories for exposition could be contrasted with theories of narrative representation to aid the analysis of similarities and differences between genres with respect to representation and corresponding process.

Expanding the scope of relations studied on-line

The vast majority of studies of on-line text processing have examined how readers relate the statement they are currently reading to relevant information in their text representations. These studies examine how readers identify or infer concept-to-concept relations and proposition-to-proposition or event-to-event relations in the course of comprehending a sentence in a text.

Perhaps because of the emphasis on study of narratives, the great majority of studies of inferential processing have focused on two categories of relations between concepts and sentences—referential relations and causal relations. There have been several studies of processing of spatial relations (Morrow *et al*., 1987; Zwaan & Oostendorp, 1993) and some study of instrumental inferences (Dosher & Corbett, 1982; Keenan & Jennings, 1995). In addition, isolated studies of processing of various other types of relations can be cited, including logical relations (Lea, 1995) and temporal relations (Trabasso, van den Broek, & Suh, 1989; Zwaan, 1996; Zwaan, Magliano, & Graesser, 1995). However, a great deal of research remains to be done on processing of relations other than anaphoric and causal relations. In this regard, it would be most useful to develop theories concerning the types of mental models that are communicated in various types of exposition. For example, what type of representation do readers construct on the basis of reading a set of instructions (e.g., how to assemble a toy)? As another example, what type of representation do readers construct from a description of a scientific theory (e.g., the particle theory of light), principle (e.g., the uncertainty principle in physics), or process (e.g., mitosis)?

Virtually all of the research examining on-line comprehension processes—including most of the work on processing of referential and causal relations—looks at how readers construct horizontal connections between pairs of concepts or events in a text. There is much less attention to the question of how readers process vertical relations in a text. There are various types of vertical relations: Some text statements have a superordinate relation to several subsequent statements (e.g., topic sentences in exposition); some sequences of statements support a generalization that subsumes them; other sequences of statements may represent an instantiation of a script that subsumes them. Although Kintsch & van Dijk (1978; van Dijk & Kintsch, 1983) discussed the importance of such relations in their theoretical statements, relatively little research has been directed toward them (but see: Cirilo & Foss, 1980; Dopkins, 1996; Kieras, 1978, 1980, 1981; Long, Golding, & Graesser, 1992; Lorch, Lorch, & Matthews, 1985; Lorch, Lorch, & Mogan, 1987; Meyer, 1975; van den Broek, 1988; van den Broek & Lorch, 1993). Given that most texts are hierarchically structured and that recall often is dominated by these vertical relations (Kintsch & van Dijk, 1978; Lorch, Lorch, & Inman, 1993; Meyer, 1975), it is important to address the question of how vertical relations are processed and represented during reading.

Redefining "on-line" processing

Among the most important contributions of cognitive science to reading research has been the development of theories of on-line processing and of methodologies to test the theories. Unfortunately, in our zeal to examine on-line processes, our criteria for what constitutes evidence of on-line processing

may have become too restrictive. The most extreme case is when probe procedures (e.g., naming, word recognition) are used to determine whether a specific concept is "activated" at a particular point during reading. These procedures have often been used to test for the on-line occurrence of an inference (Potts *et al*., 1988). In order for an inference to be detected by this procedure, several conditions must be met: (a) the probe word selected to test for an inference must, in some sense, be activated by the inference; (b) the targeted inference must be relatively highly constrained in content in order for a single probe word to be sensitive; (c) the inference must have been made at the point in time the probe word is presented, usually immediately upon reading the sentence that is hypothesized to evoke the inference; (d) all of the preceding conditions must be met by a majority of readers. The probe procedure has been very useful in demonstrating that adult readers do consistently make certain classes of inferences under predictable conditions. However, the success of the method may obscure its limitations. The procedure will produce null results under many circumstances: If different readers make different inferences; if inferences do occur, but only relatively slowly; if the target word is poorly selected or if no single target word is sufficiently sensitive to the occurrence of the target inference (Magliano & Graesser, 1991).

Other methodologies are more forgiving than probe tasks. Eye-tracking and sentence-reading procedures are sensitive to slowdowns in reading that presumably reflect extracognitive processing associated with making an inference. However, these procedures do not reveal the content of the extra processing. Also, like the probe procedures, they assume a close time-locking of inferential activity to specific, triggering information in a text. The requirement that inferential activity be closely time-locked to a specifiable point in a text probably derives from Just and Carpenter's (1980) immediacy assumption. It is impressive that cognitive science has been able to demonstrate that many inferences are made in very close temporal relation to a specifiable text event. However, it is possible that we are failing to observe many types of inferences that are made during reading, but are not necessarily made in lockstep fashion. For example, time-observing procedures will not be sensitive to an inference if the inference is made relatively slowly (e.g., two sentences after the triggering event) or if there are substantial individual differences in when the inference is made.

Recently, a procedure has been developed that avoids some of the pitfalls of word probe procedures and does not presume that an inference is made at a specifiable point in time by all readers (O'Brien & Albrecht, 1992). In this procedure, readers are presented with a text in which a target statement is inconsistent with information encountered earlier during reading. If readers monitor the global coherence of a text, then the processing of the target statement should reflect their recognition of the inconsistency. Indeed, readers slow down their reading when they encounter such inconsistencies in a text

(Albrecht & O'Brien, 1993; Myers *et al*., 1994; O'Brien & Albrecht, 1992). Of particular interest in the present context, the procedure can also be used to test whether a specific *inference* was made at some earlier point in a text. For example, if a description of the protagonist supports the inference that the protagonist is a vegetarian (without explicitly saying so), a subsequent probe statement about eating a cheeseburger would reveal whether the vegetarian inference was made. Given demonstrations that readers are sensitive to inconsistencies across relatively long surface distances in a text (Albrecht & Myers, 1995; Albrecht & O'Brien, 1993; Myers *et al*., 1994), the procedure does not entail the requirement that the making of a specific inference be closely time-locked to a specific text event. In addition, it seems more reasonable to assume that a probe consisting of a sentence will more adequately tap a target inference than a probe consisting of a single word.

The contradiction task is an example of a procedure that implicitly defines on-line processing as processing that is demonstrated to occur during reading. However, unlike probe procedures and many applications of sentence-reading and eye-tracking procedures, the contradiction task does not require a precise temporal relation between the generation of an inference and a specific text event. Perhaps other procedures may be developed that are similar in the characteristic of being less restrictive in the criteria they impose for the demonstration of on-line inferencing. Given the likelihood that many types of inferences are not triggered in response to a single, well-specified text event (e.g., vertical inferences of various types), the development of such alternative procedures should be an important goal in future research.

Strategic aspects of reading comprehension

As in many other areas of cognitive science, researchers have devoted much of their attention to the nature of the relatively automatic "microprocesses" of reading (cf. McKoon & Ratcliff, 1992). Although this focus has led to important advances, it has also led to serious omissions in the domain of text comprehension.

To begin, cognitive science has paid little attention to the general question of how readers' goals influence their reading behavior. Some theorists (e.g., McKoon & Ratcliff, 1992) seek to neutralize the influence of readers' goals in their experiments so that the automatic processes underlying all reading can best be studied. Their premise is that a set of core reading processes exists that is unaffected by reading goals. Other theorists (e.g., Graesser *et al*., 1994; Kintsch & van Dijk, 1978; Singer *et al*., 1994) assert that the purpose for which a person reads has an overarching influence on text processing and that it is not meaningful to conceptualize reading in the absence of a goal. Regardless of one's stand on this issue, the fact remains that cognitive science has directed little attention to the question of how reading goals influence text processing. Given the plausibility of the hypothesis that

readers' goals are an important determinant of their processing strategies, we strongly encourage systematic investigation of the interaction of reading goals and text processing (see van den Broek *et al*., 1993, 1995).

The starting point for research in this direction should be the identification of the distinct goals for which readers read texts (Lorch *et al*., 1993, 1995). Once a typology of reading goals is established, theorizing should address the ways in which a reader's goals might influence processing of a text (Trabasso, Suh, Payton, & Jain, 1995; van den Broek *et al*., 1995; Zwaan, 1994). Goals may influence reading by any of several mechanisms: (a) They may affect setting of gross parameters of reading behavior, such as reading speed or the extent of activation of information in long term memory; (b) they may lead to the adoption of specific text processing strategies that are cued to aspects of text structure and/or content (e.g., summarization or search); (c) they may lead to the activation of specific background knowledge to serve as a context for the interpretation of the text; (d) they may affect the criteria adopted by readers for monitoring their reading (van den Broek *et al*., 1995). Any of these potential "macro" effects of goals may have pervasive effects on reading behavior, including effects on microprocesses during reading. For example, the number and types of inferences a reader generates during reading may be influenced by any and all of the hypothesized mechanisms.

Another consequence of cognitive science's bias for studying microprocesses is that little attention has been paid to the nature of various global processing strategies that might be important in reading. This point is closely related to the topic of reading goals in that it seems likely that mature readers have repertoires of text processing strategies associated with different reading goals. In addition, readers may have global processing strategies associated with the construction of a hierarchically organized text representation (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). McKoon and Ratcliff (1992) have argued that global processing strategies are utilized only under the influence of "special" reading goals and have pointed to the lack of empirical evidence in the cognitive literature for such strategic processing. In fact, there have been relatively few attempts in the cognitive literature to study strategic processing and the great majority of empirical investigations have not examined conditions that would encourage readers to use global processing strategies: In most studies, the texts are very short and simple in structure and the tasks set for readers are generally impoverished, unfamiliar, or both (e.g., read in preparation for a word recognition test).

Finally, cognitive scientists have been somewhat restricted in their consideration of individual differences in text comprehension skills. Typically, cognitive researchers have examined how individual differences in basic cognitive abilities (e.g., working memory capacity) covary with performance on measures of text processing or comprehension (Daneman & Carpenter,

1980, 1983; Just & Carpenter, 1992; Whitney, Ritchie, & Clark, 1991). This is certainly a sensible approach as many differences in readers' text processing abilities may be due to differences in more basic cognitive abilities or resources. However, it is also possible that an entire category of individual differences is being overlooked. Namely, readers may develop different processing strategies—particularly at level of various types of macrostrategies—not because of differences in basic cognitive abilities, but simply as a result of learning experiences or motivation. As more attention is directed to strategic aspects of text processing, we might expect to find more individual variation in processing.

There are at least two ways in which to approach the study of individual differences in processing strategies. One is to attempt to anticipate the variety of strategies used by readers, as well as the types of individual difference variables that might correlate with the use of different processing strategies. This top-down approach presumes quite a bit of knowledge of both reading strategies and individual differences in order to be successful. A complementary, bottom-up approach would be to develop empirical strategies for trying to induce the text processing strategies of different readers in different reading situations. For example, reading behavior might be assessed with a battery of on-line processing measures designed to provide both a comprehensive picture of text processing and to contain indicators of potential ways in which strategies may differ. Performance across the set of measures might be analyzed by inductive statistical procedures (e.g., cluster analyses) to detect potential systematic individual differences in text processing strategies. In domains where basic knowledge about the nature of strategic processing is lacking, this bottom-up approach could supply the necessary information to guide a top-down strategy for researching individual differences in text processing strategies.

Exploring the upper limits of readers' abilities

Cognitive science has focused almost exclusively on the question of what readers do in laboratory reading tasks, rather than on question of what readers *can* do. Unfortunately, most cognitive experiments investigate reading under conditions that do not encourage very extensive comprehension processing. Participants in experiments may read 20 to 40 brief texts in a 50-min session with a simple comprehension question presented after each text, or after every few texts, to force the "reading for meaning." The emphasis is on-line measures of processing and generalizability of findings over stimulus materials. However, the impoverished reading conditions may create a self-fulfilling prophecy: Cognitive scientists establish "data limited" situations for studying reading, then interpret their findings as reflecting "resource limits" on the part of the reader (Norman & Bobrow, 1975). That is, the minimal processing exhibited by readers may be attributable to constraints

imposed by the *reading conditions* under which they are typically observed, rather than reflecting the capabilities of the *readers*.

It is important to develop methodologies that push readers to the upper limits of their abilities to comprehend a text. This suggestion reinforces many of the suggestions made in previous sections of this paper. We should make greater use of experimental procedures, such as think aloud, that encourage extensive text processing (Pressley & Afflerbach, 1995; Suh & Trabasso, 1993; Trabasso & Magliano, 1996). We should study processing of texts that are longer and more complexly structured than those typically investigated in cognitive research. We should systematically manipulate readers' goals. In short, we should study reading under conditions that encourage readers to demonstrate their intelligence.

Conclusion

Cognitive science has made impressive contributions to our understanding of reading comprehension in the past 2 decades. Armed with a variety of techniques to examine readers' moment-to-moment comprehension efforts as they read, cognitive scientists have provided detailed descriptions of inferential processes that are central reading. Conceptually, they have provided a definition of comprehension and a framework for its study that should continue to be very fruitful. Our hope for the future is that cognitive science will expand its theories and methodologies to address the full range of readers' abilities and experiences.

Acknowledgments

We thank Jose Leon and Mike Royer for their helpful comments on the initial draft of the manuscript.

References

Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1459–1468.

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 1061–1070.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington: Winston.

Anderson, R. C., & Myrow, D. L. (1971). Retroactive inhibition of meaningful discourse. *Journal of Educational Psychology*, **62**, 81–94.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal learning. *Journal of Educational Psychology*, **51**, 267–272.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge Univ. Press.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Cambridge, MA: Blackwell.

Bloom, C. P., Fletcher, C. R., van den Broek, P., Reitz, L., & Shapiro, B. P. (1990). An on-line assessment of causal reasoning during text comprehension. *Memory & Cognition*, **18**, 65–71.

Bransford, J. D., Barclay, J., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, **3**, 193–209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, **2**, 331–350.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, **11**, 717–726.

Brewer, W. F. (1980). Literary theory, rhetoric, and stylisties: Implications for psychology. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 221–239). Hillsdale, NJ: Erlbaum.

Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, **83**, 329–245.

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 1–30). Washington: Winston.

Cirilo, R. K., & Foss, D. J. (1980). Text structure and reading time for sentences. *Journal of Verbal Learning and Verbal Behavior*, **19**, 96–109.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, **82**, 407–428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 240–247.

Crouse, J. H. (1971). Retroactive interference in reading prose materials. *Journal of Educational Psychology*, **62**, 39–44.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, **19**, 450–466.

Daneman, M., & Carpenter, P. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 561–584.

Dee-Lucas, D., & Larkin, J. H. (1988). Attentional strategies for studying scientific text. *Memory & Cognition*, **16**, 469–479.

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, **88**, 216–222.

Dopkins, S. (1996). Representation of superordinate goal inferences in memory. *Discourse Processes*, **21**, 85–104.

Dosher, B. A., & Corbett, A. T. (1982). Instrument inferences and verb schemata. *Memory & Cognition*, **10**, 531–539.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, **22**, 75–87.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211–245.

Fletcher, C. R. (1981). Short-term memory processes in text comprehension. *Journal of Verbal Learning and Verbal Behavior*, **20**, 564–574.

Fletcher, C. R. (1986). Strategies for the allocation of short-term memory during comprehension. *Journal of Memory and Language*, **25**, 43–58.

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, **27**, 235–244.

Frase, L. T. (1968). Effect of question location, pacing, and mode upon retention of prose materials. *Journal of Educational Psychology*, **60**, 49–55.

Frase, L. T. (1969). Cybernetic control of memory while reading connected discourse. *Journal of Educational Psychology*, **60**, 49–55.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, **14**, 178–210.

Frederiksen, C. H. (1975). Acquisition of semantic information from discourse: Effects of repeated exposures. *Journal of Verbal Learning and Verbal Behavior*, **14**, 158–169.

Garrod, S., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 250–257.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. W. (1992). Do readers mentally represent characters' emotional states? *Cognition and Emotion*, **6**, 89–111.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1–67.

Goldman, S. R., & Varma, S., & Cote, N. (1995). CAPping the construction-integration model of discourse comprehension. In C. A. Weaver, S. M. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337–358). Hillsdale, NJ: Erlbaum.

Goldman, S. R., & Varnhagen, C. K. (1986). Memory for embedded and sequential story structures. *Journal of Memory and Language*, **25**, 401–418.

Graesser, A. C., Bertus, E. L., & Magliano, J. P. (1995). Inference generation during the comprehension of narrative text. In R. F. Lorch, Jr. & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 295–320). Hillsdale, NJ: Erlbaum.

Graesser, A. C., & Clark, L. F. (1985). *The structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101**, 371–395.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, **13**, 512–521.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, **60**, 549–571.

Holyoak, K. J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. A. Eriesson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 301–335). London: Cambridge Univ. Press.

Johnson, R. E. (1972). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior*, **9**, 12–20.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, **87**, 329–354.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122–149.

Kamas, E. N., & Reder, L. M. (1995). The role of familiarity in cognitive processing. In R. F. Lorch. Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 177–202). Hillsdale, NJ: Erlbaum.

Keenan, J. M., & Jennings, T. M. (1995). The role of word-based priming in inference research. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 37–50). Hillsdale, NJ: Erlbaum.

Kieras, D. E. (1978). Good and bad structure in simple paragraphs: Effects on apparent theme, reading time, and recall. *Journal of Verbal Learning and Verbal Behavior*, **17**, 13–28.

Kieras, D. E. (1980). Component processes in the comprehension of simple prose. *Journal of Verbal Learning and Verbal Behavior*, **20**, 1–23.

Kieras, D. E. (1981). Topicalization effects in cued recall of technical prose. *Memory & Cognition*, **9**, 541–549.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, **95**, 163–182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, **85**, 363–394.

Klin, C. M. (1995). Causal inferences in reading: From immediate activation to long-term memory. *Journal of Experimental Psychology: Learning. Memory, and Cognition*, **21**, 1483–1494.

Klin, C. M., & Myers, J. L. (1993). Reinstatement of causal information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 554–560.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.

Lea, R. B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1469–1482.

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). The generation of goal related inferences during narrative comprehension. *Journal of Memory and Language*, **5**, 634–647.

Lorch, R. F., Jr., Klusewitz, M. A., & Lorch, E. P. (1995). Distinctions among reading situations. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 375–398). Hillsdale, NJ: Erlbaum.

Lorch, R. F., Jr., Lorch, E. P., & Inman, W. E. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology*, **85**, 281–290.

Lorch, R. F., Jr., Lorch, E. P., & Klusewitz, M. A. (1993). College students' conditional knowledge of about reading. *Journal of Educational Psychology*, **85**, 239–252.

Lorch, R. F., Jr., Lorch, E. P., & Matthews, P. D. (1985). On-line processing of the topic structure of a text. *Journal of Memory and Language*, **24**, 350–362.

Lorch, R. F., Jr., Lorch, E. P., & Mogan, A. M. (1987). Task effects and individual differences in on-line processing of the topic structure of a text. *Discourse Processes*, **10**, 63–80.

Magliano, J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, **20**, 193–232.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, **17**, 578–586.

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, **25**, 645–656.

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning. Memory, and Cognition*, **22**, 919–932.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning. Memory, and Cognition*, **12**, 82–91.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, **99**, 440–466.

McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 1219–1228.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North-Holland.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard Univ. Press.

Miller, J. R., & Kintsch, W. (1980). Readability and recall for short passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 335–354.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, **26**, 165–187.

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 876–886.

Myrow, D. L., & Anderson, R. C. (1972). Retroactive inhibition of prose as a function of type of test. *Journal of Educational Psychology*, **68**, 303–308.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7, 44–64.

Norman, D. A., & Rumelhart, D. E. (1975). *Explorations in cognition*. San Francisco: Freeman.

O'Brien, E. J. (1995). Automatic components of discourse comprehension. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 159–176). Hillsdale, NJ: Erlbaum.

O'Brien, E. J., & Albrecht, J. E. (1991). The role of context in accessing antecedents in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 94–102.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 777–784.

O'Brien, E. J., & Myers, J. L. (1987). The role of causal connection in the retrieval of text. *Memory & Cognition*, **15**, 419–427.

O'Brien, E. J., Plewes, P. S., & Albrecht, J. E. (1990). Antecedent retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 241–249.

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 410–420.

Pompi, K. E., & Lachman, R. (1967). Surrogate processes in the short-term retention of connected discourse. *Journal of Experimental Psychology*, **75**, 143–150.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, **27**, 399–415.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of the constructively responsive reader*. Hillsdale, NJ: Erlbaum.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59–108.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, **7**, 65–81.

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology*, **39A**, 657–673.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, **16**, 829–837.

Rothkopf, E. Z. (1966). Learning from written materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, **3**, 241–249.

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, **58**, 56–61.

Royer, J. M. (1977). Remembering: Constructive or reconstructive? Comments on Chapter 5 by Spiro. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 167–174). Hillsdale, NJ: Erlbaum.

Royer, J. M., Carlo, M. S., Dufresne, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and Instruction*. **14**, 373–408.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, **2**, 437–442.

Sanford, A. J., Garrod, S., & Boyle, J. M. (1977). An independence of mechanism in the origins of reading and classification-related semantic distance effects. *Memory & Cognition*, **5**, 214–220.

Schank, R. (1975). *Conceptual information processing*. Amsterdam: North-Holland.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, **25**, 279–294.

Shirey, L. S., & Reynolds, R. E. (1988). Effect of interest on attention and learning. *Journal of Educational Psychology*, **80**, 159–166.

Singer, M. (1980). The role of case-filling inferences in the coherence of brief passages. *Discourse Processes*, **3**, 185–201.

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, **22**, 437–448.

Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language*, **33**, 421–441.

Spiro, R. J. (1977). Remembering information from text: The "state of schema" approach. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 137–166). Hillsdale, NJ: Erlbaum.

Stein, N. L., & Levine, L. (1990). Making sense out of emotional experience: The representation and use of goal-directed knowledge. In N. L. Stein, B. Leventhal, & T. Trabasso (Eds.), *Psychological and biological approaches to emotion* (pp. 45–73). Hillsdale, NJ: Erlbaum.

Stein, N. L., Liwag, M. D., & Wade, E. (1995). A goal-based approach to memory for emotional events: Implications for theories of understanding and socialization. In R. D. Kavanaugh, B. Z. Glick, & S. Fein (Eds.), *Emotion: Interdisciplinary approaches* (pp. 91–118). Hillsdale, NJ: Erlbaum.

St. John, M. F. (1992). The story gestalt: A model of knowledge intensive processes in text comprehension. *Cognitive Science*, **16**, 271–306.

Suh, S., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language*, **32**, 279–300.

Sulin, R. A., & Dooling, D. J. (1974). Intrusions of a thematic idea in retention of prose. *Journal of Experimental Psychology*, **103**, 255–262.

Trabasso, T., & Magliano, J. P. (1995). Understanding emotional understanding. In N. Frijda (Ed.), *The Proceedings of the International Society of Research on Emotions* (pp. 78–82). Hillsdale, NJ: Erlbaum.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes*, **21**, 255–288.

Trabasso, T., Secco, T., & van den Broek, P. (1981). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Erlbaum.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, **24**, 595–611.

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes*, **16**, 3–34.

Trabasso, T., Suh, S., Payton, P., & Jain, R. (1995). Explanatory inferences and other strategies during comprehension and their effect on recall. In R. E. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 219–240). Hillsdale, NJ: Erlbaum.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, **24**, 612–630.

Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal relations in the representation of stories. *Discourse Processes*, **12**, 1–25.

van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, **27**, 1–22.

van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423–446). Hillsdale, NJ: Erlbaum.

van den Broek, P., Fletcher, C. R., & Risden, K. (1993). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes*, **16**, 169–180.

van den Broek, P., Lorch, E. P., & Thurlow, R. (in press). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories and hierarchical level. *Child Development*.

van den Broek, P., & Lorch, R. F., Jr. (1993). Network representations of causal relations in memory for narrative texts. *Discourse Processes*, **16**, 75–98.

van den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 165–187). Mahwah, NJ: Erlbaum.

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch, Jr. & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–374). Hillsdale, NJ: Erlbaum.

van den Broek, P., Rohleder, L., & Navarez, D. (1996). Causal inferences in the comprehension of literary texts. In R. J. Kreuz & M. S. MacNealy (Eds.), *Empirical approaches to literature and aesthetics* (pp. 179–200). Norwood, NJ: Ablex.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Orlando, FL: Academic Press.

Vonk, W., & Noordman, L. G. (1990). On the control of inferences in text understanding. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 447–463). Hillsdale, NJ: Erlbaum.

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, **14**, 133–145.

Zwaan, R. A. (1994). Effects of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 920–933.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, 1196–1207.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 386–397.

Zwaan, R. A., & van Oostendorp, H. (1993). Do readers construct spatial representations in naturalistic story comprehension? *Discourse Processes*, **16**, 125–143.

Part XIII

MATHEMATICS

69

DEVELOPING MATHEMATICAL KNOWLEDGE

L. B. Resnick

Recent research has led to a significant reconceptualization of the nature of children's number knowledge development. This article outlines infants' and preschoolers' implicit protoquantitative reasoning schemas and shows how these combine with early counting knowledge to produce mathematical concepts of number. Research on elementary school children's informal and invented arithmetic is reviewed, and implications for mathematics education are evaluated.

Research on children's knowledge and learning of mathematics has been one of the most active topics in developmental cognitive psychology in recent years. The result has been not only an explosion of research studies but also a significant reconceptualization of the nature of early mathematical knowledge, of how children acquire such knowledge informally, and of how mathematics learning proceeds in school. Relevant research has been conducted by cognitive, developmental, and educational psychologists as well as by a vibrant community of mathematics educators. Despite their diverse training and affiliation, there is broad agreement among these various research groups on what can be termed a *constructivist* assumption about how mathematics is learned. It is assumed that mathematical knowledge—like all knowledge—is not directly absorbed but is constructed by each individual. This constructivist view is consonant with the theory of Jean Piaget but comes in many varieties and does not necessarily imply either a stage theory or the logical determinism of orthodox Piagetian theory.

Psychologists today attribute much more mathematical knowledge and understanding to children than they once did. One reason for this is that they have learned to look beyond what children say explicitly for knowledge that is implicit in what they do. In current research on children's understanding of mathematics, for example, psychologists look at the

Source: *American Psychologist*, 1989, 44(2), 162–169.

invented procedures children develop as one kind of evidence of their implicit knowledge. When children demonstrate how they do an arithmetic problem, for example, researchers try to determine what implicit rules they may be following. They are willing to grant children implicit knowledge of the principles instantiated by the rules if the rules are followed regularly, even when children are completely unable to articulate them. (See Gelman & Greeno, in press, for a discussion of the nature of implicit principles in children's mathematics.)

This article reviews what is now known about the nature of children's mathematics knowledge and learning. It concentrates on knowledge of numbers and, thus, a counting and arithmetic, both because this is where the most research has been done and because numbers and arithmetic form the core of the elementary and middle-school curriculum. There seems to be general consensus that number concepts form the basis upon which higher mathematical competencies can develop. As I will show, however, there is less consensus about the best ways of teaching number concepts and arithmetic. As I review the research evidence on number concept development I will also consider what is known about alternative instructional approaches. Then, in conclusion, I will examine the implications of the research findings as a whole for general concerns that have been expressed in society about mathematical development.

Quantity concepts in the preschool years

Infants' preverbal quantitative knowledge

The process of constructing mathematics knowledge begins well before school. Some of the basic elements of quantity knowledge appear to be present in babies well before they could have been taught even informally. Infants of about 6 months can discriminate the numerosity of small sets when these are presented visually. What is more, they can match sets cross-modally, recognizing the same quantity whether it is presented visually or auditorily (Starkey, Spelke, & Gelman, in press). This research shows that babies already know about units and repetitions, and that, within a limited range, they are able to recognize exact differences in how many units or repetitions they have experienced. Babies also know about size differences. Infant discrimination studies show that they are able to make judgments on the basis of comparative rather that absolute size. This means that they have some kind of schema for comparing objects quantitatively. Babies' quantitative knowledge is, of course, prelinguistic. As language develops, two additional kinds of knowledge become available: a variety of protoquantitative terms and concepts that express quantity without numerical precision; and numerical quantification, with counting as its primary mechanism and expression.

Protoquantitative schemas

During the preschool years, children develop a large store of nonnumerical quantity knowledge. They express quantity judgments in the form of absolute size labels such as *big*, *small*, *lots*, and *little*. They also begin to put linguistic labels on the comparisons of sizes they made as infants. Thus, they can look at two circles and declare one bigger than the other, see two trees and declare one taller than the other, examine two glasses of milk and declare that one contains more than the other. It is useful to think of these judgments as based on a protoquantitative comparison schema, one that operates perceptually, without any measurement process.

Two additional protoquantitative schemas can be identified. One interprets changes as increases or decreases in quantities. This protoquantitative increase/decrease schema allows children as young as 3 or 4 years of age to reason that, if they have a certain amount of something, and they get another amount of the same thing (perhaps mother adds another cookie to the two already on the child's plate), they have more than before. Or, if some of the original quantity is taken away, they have less than before. Equally important, children know that, if nothing has been added or taken away, they have the same amount as before. For example, children show surprise and label as "magic" any change in the number of objects on a plate that occurs out of their sight. This shows that children have the underpinnings of number conservation well before they can pass the standard Piagetian (Piaget, 1941/1965) tests. They can be fooled by perceptual cues or language that distracts them from quantity, but they possess a basic understanding of addition, subtraction, and conservation.

As a result of their everyday experience, preschool children also have a protoquantitative part–whole schema. They know about the ways in which material around them comes apart and goes together, that it is *additive*. That is, one can cut a quantity into pieces that, taken together, equal the original quantity. One can put two quantities together to make a bigger quantity and then join that bigger quantity with yet another in a form of hierarchical additivity. In a nonlinguistic, implicit way, children know about this additive property of quantities. This protoquantitative knowledge allows them to make judgments about the relations between parts and wholes. Children know, for example, that a whole cake is bigger than any of its pieces. They can make this judgment logically without needing to actually see the cake and its parts. In fact, visual inspection is as likely to lead to misjudgments as to support correct ones.

This claim that children understand part–whole relations might sound contradictory to the well-established Piagetian findings concerning preschool children's inability to reliably solve class-inclusion problems, such as "Which is more—the brown beads or the brown and the white beads together?" (Inhelder & Piaget, 1964). Studies have now shown, however, that if labels

are used to focus the child's attention clearly on the whole collection rather than its individual members (e.g., speaking of a forest instead of pine trees plus oak trees), children as young as 4 or 5 years make correct class-inclusion judgments. Thus one can attribute to preschool children an implicit set of principles about part–whole relations and additive composition.

In past accounts of children's cognitive development, especially in the Piagetian tradition, much has been made of the limits of preschoolers' quantity knowledge. It is true that preschoolers' protoquantitative knowledge lacks certain basic measurement rules. Preschoolers do not typically know, for example, that to compare the lengths of two sticks it is necessary to align them at one of the ends. Furthermore, if a row of objects *looks* longer, children of this age will usually declare it to have more objects. This is the essence of the preschoolers' tendency not to conserve quantity under perceptual transformations. Although these limits on preschoolers' knowledge are real, it is just as reasonable, and perhaps more useful, to focus attention on what very young children *do* know about quantity. Preschool children clearly have a tendency to respond to perceptually apparent differences that do not always correlate with true quantity differences. They also have some difficulty with the language of sets and subsets. Experiments such as the "magic" and the collection labeling studies, however, show that young children already understand implicitly some of the basic logic of quantity relations. Children's protoquantitative reasoning schemas constitute a major foundation for later mathematical development. When eventually coordinated with counting skill, which develops separately during the preschool years, they form the basis for understanding several major principles of the number system.

Counting and exact quantification

Counting, a culturally transmitted formal system, is the first step in making quantitative judgments exact. It is a measurement system for sets. Gelman and her colleagues have done the seminal work analyzing what it means to "understand" counting, showing that children as young as three or four years of age implicitly know the key principles that allow counting to serve as a vehicle of quantification (Gelman & Gallistel, 1978). These principles include the knowledge that number names must be matched one-for-one with the objects in a set and that the order of the number names matters, but the order in which the objects are touched does not. Knowledge of these principles is inferred from the ways in which children solve novel counting problems. For example, if asked to make the second object in a row "number 1," children do not neglect the first object entirely but, rather, assign it one of the higher number names in the sequence.

Other research has challenged Gelman's assessment of the ages at which children can be said to have acquired all of the counting principles. Some of the challenges are really arguments about the criteria for applying certain

terms. For example, Gelman has attributed knowledge of *cardinality*, a key mathematical principle, to children as soon as they know that the last number in a counting sequence names the quantity in the whole set; others would reserve the term for a more advanced stage in which children reliably conserve quantity under perceptual transformations. A challenge that goes beyond matters of terminology comes from research showing that, although children may know all of the principles of counting and be able to use counting to quantify given sets of objects or to create sets of specified sizes, they may not, at a certain stage, have fully integrated their counting knowledge with their protoquantitative knowledge. Sophian (1987), for example, has shown that children who know how to count sets when directly asked, "How many are there?" do not spontaneously count when asked to solve conservation and similar problems. Gelman and Greeno (in press) have suggested that a distinction between knowledge of principles and knowledge of how to interpret particular situational demands can explain such findings, allowing one to attribute knowledge of the principles to children even when they do not apply them in all possible situations.

Either way, such findings make it clear that, even after knowledge of counting principles is established, there is substantially more growth in number concepts still to be attained. A first major step in this growth is integration of the number-name sequence with the protoquantitative comparison schema. This seems to happen as young as about four years of age. At this point, children behave as if the counting word sequence constitutes a kind of "mental number line" (Resnick, 1983). They can quickly identify which of a pair of numbers is "more" by mentally consulting this number line, without actually stepping through the sequence to determine which number comes later. In the child's subsequent development, counting as a means of quantifying sets is integrated with the protoquantitative part–whole and increase–decrease schemas. This is when stable class inclusion and conservation performances will appear. When numerical quantity becomes the dominant way in which children think of quantity, they will not be driven by perceptual and linguistic cues.

It is not clear exactly what role instruction—whether formal in the preschool or informal within the family—plays in this development. There is good evidence that the kind of basic quantitative knowledge discussed here —counting and its use in connection with set combination and increase–decrease situations—is universal, although it apparently develops at different rates in different cultures. Different family subcultures no doubt influence the rate of development as well. Although there are no supporting controlled studies, it seems very probable that the most important difference among cultures and subcultures in this respect is the extent to which quantification is a frequent, everyday occurrence. In environments in which exact quantification is frequently demanded, children are likely to pay more attention to numbers, to learn the number sequence sooner, and to count sets sooner. As

a result, they will probably also quantify combination and increase–decrease situations sooner. If this is so, preschool programs can foster number-concept development mainly by providing many occasions and requests for quantification and by eventually tying these requests to situations of comparison, combination, and increase or decrease of quantities. Television programs such as *Sesame Street*, in which numbers and counting are highlighted, probably also produce earlies number development. Both kinds of intervention will produce their greatest effect when they stimulate lots a counting in unsupervised everyday situations. Only in this way is enough practice likely to be generated to produce a real effect on the rate of number-concept development.

Conceptual development in the elementary school years

It is much more difficult to tell a definite story about development of mathematical concepts during the elementary school years. This is partly because the range of mathematical concepts to be learned becomes much broader, and only a few of these concepts have been intensively studied. It is also because school instruction has strong, but not completely understood, effects on development. In an attempt to limit the complexity, I will begin by looking for aspects of conceptual development that appear to be independent of specific schooling practice. Then, in light of evidence about what is perhaps universalt in mathematics development, I will examine schooling practices and evaluate their probable effects.

Invented strategies for calculation

One way to establish the nature of children's developing number concepts is to examine some of their invented informal strategies for doing arithmetic. Children's strategies for adding and subtracting have been well documented in multiple studies in many countries (see Resnick, 1983, for a review). Most children use their knowledge of counting to calculate answers. At first they count actual objects—most often their fingers—thus directly modeling the way in which the material of the world combines and decomposes. For addition, they create two sets of objects, one for each of the addends, combine the two sets, and recount the newly combined set. For subtraction, they count out a "starting set," remove the specified number of objects from the set, and then recount the remainder. Some children are adept at such methods when they enter school. Virtually all children develop them eventually.

For most children, extensive practice in counting actual objects eventually leads to an ability to use the count words ("one," "two," "three," etc.) themselves as objects to count, and they thus become able to engage in mental counting. This produces both efficiency and flexibility in solving addition and subtraction problems. It also gives evidence of considerable mathematical understanding. By age six or seven, for example, most children invent a

way of mentally computing addition problems that minimizes the amount of counting they must do. Their method has come to be called *counting on*. To add two numbers, children behave as if they are setting a mental counter-in-the-head to one of the addends, and then count on by ones enough times to "add in" the second addend. Thus, to add 5 and 3, children might say to themselves, "5 . . . 6, 7, 8," giving the final count word as the answer. What is more, children do not always start with the first number given in a problem but will invert the addends to minimize the number of counts when necessary. Thus, in adding 3 + 5, they perform exactly the same procedure as for adding 5 + 3. Children's willingness, in a procedure they invent for themselves, to count *on*—without first counting *up* to the first number—demonstrates that they have come to appreciate that "a 5 is a 5 is a 5. . . ." They know that number is not elastic, and, therefore, a count to 5 always involves the same number of objects. In addition, children's willingness to invert the addends shows that they implicitly appreciate the mathematical principle of commutativity of addition. It will be some time, however, before they will show knowledge of commutativity in a general way, across situations, across numbers, and, above all, with an ability to *talk about* rather than just *apply* the principle.

Other invented counting-based strategies used by children give further evidence that they are developing implicit knowledge of number principles, even when these are not directly taught in school. By about 9 years of age, children compute subtraction problems by either counting up from the smaller number or counting down from the larger, whichever requires fewer counts. So for the problem 9 – 2, children will say, "9 . . . 8, 7. . . . The answer is 7." But for the problem 9 – 7, a child will say, "7 . . . 8, 9. . . . The answer is 2." The first procedure is straightforward and corresponds to a very basic way of thinking about subtraction—taking away. But the second procedure actually converts a subtraction problem into a special form of addition—addition with a missing addend.

How do children know that this is permissible? Their willingness to do such conversion must be grounded in their knowledge, implicit though it may be, of the complementarity of addition and subtraction. This complementarity, in turn, is grounded in a basic principle of number, the principle of *additive composition*, which says numbers are composed of other numbers, and any number can be decomposed into parts. In children's invented counting procedures for subtraction, then, there is evidence that they have linked their counting knowledge to their protoquantitative part–whole schema, which was the preschooler's version of additive composition.

Story problems

Another source of evidence for conceptual development that appears under many different forms of schooling comes from research on how children

solve arithmetic story problems. These problems describe situations in which quantities are manipulated or evaluated. They are used in the curriculum with the intention of linking practice in arithmetic calculation to situations in which arithmetic might be applied. The extent to which the kinds of problems currently used in school meet this goal is disputed. Whatever their direct instructional value, however, story problems provide a window on children's conceptual interpretations of arithmetic.

Research from various countries (see Riley & Greeno, 1988, for a review and analysis) has shown that different classes of situations—that is, different story semantics—pose differential difficulty for children. The different semantic classes correspond closely to the kinds of protoquantitative schemas children develop during the preschool years; that is, they involve changes (increases or decreases) in a single starting quantity, or combinations of two quantities, or comparisons of two quantities.

Change stories are quite easy for children to understand when the quantity they are asked to find is the result of the increase or decrease. In such problems the solution is directly modeled by the situation. That is, if the situation describes a decrease in an initial amount, subtraction (perhaps by counting backward) of the change amount from the initial amount correctly solves the problem. When the unknown is the starting amount, however, as in the following story, the problem is much harder for children to solve:

> Ana went shopping. She spent $3.50 and then counted her money when she got home. She had $2.35 left. How much did Ana have when she started out?

Few children master these problems before eight or nine years of age. This is because there is no direct mapping between the story situation as stated and the arithmetic operation that solves the problem efficiently. The efficient arithmetic operation for solving this problem is addition (of $3.50 and $2.35), even though the story describes a decrease in the original quantity. Children eventually learn to reinterpret this kind of story situation as one that involves partitioning the initial amount into a part that is spent and a part that is retained and then combining the parts. This part–whole interpretation allows them to use their knowledge of the additive composition of numbers to reinterpret the problem in terms of an appropriate arithmetic operation. This case illustrates the complex interplay between understanding the situation and choosing the correct arithmetic operation for story problems. Additional complexity comes from the need for special forms of linguistic interpretation in order to use the text of story problems to construct a mental representation of the situation (Kintsch & Greeno, 1985).

The second major category of semantic situations describes the combining of two static quantities under a new superordinate category. For example, a certain number of apples and another number of oranges can be combined

to produce a quantity of fruit. Because they contain no real event, but rather rename the same quantities under new categories, *combine* situations have been found slightly more difficult for younger children to solve than *change* situations with the final quantity unknown.

The third category of problems describes a situation in which two quantities are compared and a numerical difference between them is found. Although deciding which set is larger is very easy for children, putting an exact numerical value on this difference is very difficult; many cannot solve this class of problems until late in elementary school. This seems to be because the difference describes a relationship between two other quantities; it is a kind of second-order concept rather than a direct quantification of material in the world.

Research on story problems involving multiplication, division, proportion, and ratio is far less developed, but there is now sufficient research to allow a distinction among several different semantic categories (Nesher, 1988). Children first understand multiplication as a kind of repeated addition ("5 books + 5 books + 5 books . . ."), and problems that can be interpreted in this way are the ones they solve earliest. Other kinds of multiplication problems (e.g., "Judy has 3 shirts and 2 skirts. For how many days will she be able to go to school wearing different outfits of a shirt and a skirt?") are much more difficult. Even children of middle-school age find such problems difficult to interpret. The same is true for many kinds of proportion and ratio problems. For all of these problems, however, as for the addition and subtraction problems, there is considerable regularity to the kinds of solutions children give at younger ages (see especially Siegler, 1981). Further analysis may reveal the existence of protoquantitative precursors to these more complex mathematical concepts, just as it already has for addition and subtraction.

Arithmetic without school: cross-cultural evidence

Children's invented arithmetic procedures show that they are able to construct basic principles of mathematics—such as commutativity, complementarity of addition and subtraction, and associativity—in intuitive forms well before such ideas are presented in school. The pattern of their performance on story problems shows that they are heavily influenced in their mathematics development by everyday experience with quantities. Research on the nature of mathematical understanding among people with little or no schooling shows a similar pattern.

A number of investigators have documented procedures that have been developed by unschooled or minimally schooled people in African, Latin American, and other developing countries. (See Resnick, 1986, for a review, and Saxe, Guberman, & Gearhart, 1987, for an important recent study.) Inspection of the procedures that are described by the researchers suggests

that procedures based on additive composition predominate. A study of children who work as street vendors in Recife, Brazil, for example, found them converting to addition, problems such as the following, which we might treat in terms of multiplication:

> How much is one coconut?
>
> 35.
>
> I'd like ten. How much is that?
>
> [Pause] Three will be 105; with three more, that will be 210. [Pause] I need four more. That is . . . [Pause] 315. . . . I think it is 350. (Carraher, Carraher, & Schliemann, 1985, p. 23).

The oral counting words of various cultures give further evidence of the universality of additive composition. Virtually all of these use some kind of a base system (usually 10) in which large numbers are denoted by combinations of smaller ones. In the Gola and Vai languages in Liberia, for example, 70 is expressed as three 20s plus 10. In each of these cultures, procedures that are efficient for the tasks common in that culture have been developed. There seems, in sum, to be strong evidence that ability to solve flexibly many kinds of additive problems develops nearly universally. The procedures used by unschooled people reveal an understanding of additive composition and principles such as commutativity, associativity, and complementarity. In contrast, ability to reason in terms of Cartesian multiplication or in terms of proportion and ratio is not so prevalent and seems to depend on very particular work experiences.

Semantics and syntax: the problem of school mathematics

Despite early and, perhaps, universal mastery of certain fundamental mathematical ideas, many children have a great deal of difficulty learning school mathematics. The phenomenon of "math anxiety"—extreme lack of confidence in one's ability to cope with mathematics—is familiar in virtually all highly educated societies. Why should strong and reliable intuitions of the kind that have been documented for young children and unschooled people not be sustained in school mathematics learning? One hypothesis with considerable empirical support is that the focus in school mathematics on formal symbol manipulation discourages children from bringing their developed intuitions to bear on school learning tasks.

A recurrent finding is that children who are having difficulty with arithmetic often use systematic routines that produce wrong answers. (See Resnick, 1987, for a review and interpretation.) This observation has been made repeatedly over the years by investigators of mathematics learning, and various studies

have documented the most common errors. A key feature of these systematic errors is that the written results tend to "look right" and to obey a large number of the important rules for manipulating symbols in written calculation. But they often disobey basic rules of how quantities behave. In subtracting from 803, for example, a child might borrow 100 from the hundreds column but return only 10 to the units column, producing a notation, $^{7}\not{8}\ 0^{1}3$. It appears that children are attending to the symbols of arithmetic but not to the quantities that they represent. School arithmetic learning, in other words, is not building effectively on the base of children's informal knowledge. The problem continues and deepens at higher levels of mathematics. Algebra students, for example, make systematic errors that show they are not thinking about the meaning—the semantics—of algebra expressions but only about the syntax, the rules for manipulating algebraic symbols. These demonstrations of syntactic performances without reference to the semantics of quantity highlight a major problem in school mathematics learning.

It is not difficult to see why children would focus on the syntax of mathematics rather than its semantics, given the way American school instruction is typically organized. In examining school textbooks and curricula for the the primary grades, one finds a sequence in which written numbers and calculations are the primary concern. New topics are usually introduced with some discussion of basic concepts, often using pictorial displays to illustrate them. But emphasis quickly shifts to practice in memorizing the basic arithmetic combinations and rules for applying these in longer multidigit computations. Memorization and written computation are stressed. Children's informal calculation methods involving counting are often suppressed very early, and there is very little oral arithmetic or extended discussion of why the taught procedures work. This kind of practice takes up most classroom time, and written computation figures very heavily in the various tests that children must take regularly.

The role and form of arithmetic practice

Although there is substantial agreement among researchers and mathematics educators about what the problem is, debate continues over how to solve it. Some believe that arithmetic practice should be sharply reduced or abandoned in favor of more conceptually oriented teaching that focuses on mathematical principles. Others claim that only a firm foundation in basic number facts and relationships will allow children to move ahead in mathematics. Two very different arguments offered in support of the latter claim lead to two distinct proposals for how to organize basic number teaching.

One argument stresses the importance of *automaticity* in retrieval of the basic number facts. If children can quickly access the basic facts used in more complex computations and mathematics problems, it is argued, their attentional resources can be devoted to remembering and performing more

complex procedures or working out new problem solutions. The argument for automaticity in arithmetic is often made by analogy to automaticity in reading, where lack of fast-access word recognition has been demonstrated to interfere with reading comprehension. A few studies specifically on arithmetic also support the claim.

The other argument treats basic arithmetic facts as the foundation for what is sometimes called "number sense." According to this argument, the number facts constitute a kind of basic structure for the number system; knowing that 3 and 2 combine to make 5 and that 5 decomposes to make two parts, 3 and 2, is a way to understand a defining feature of measure numbers—that they are additive. According to this view, a mathematically important feature of the number 5 is its decompositions, and learning basic addition and subtraction facts is likely to lead to a sense of the smaller numbers and their relations to one another. A similar sense of larger numbers and their relations could be partially acquired by learning multiplicative relations among numbers—the multiplication tables—along with learning the conventions and structures of place-value notation.

Those who argue for automaticity stress speed as well as accuracy of retrieval and often advocate substantial drill on the number facts. Supporters of the number sense point of view also value learning of basic arithmetic facts, but they are not so certain that speed of access is important. They usually propose, instead, instruction that stresses the different ways of decomposing and recomposing numbers, sometimes using oral arithmetic rather than written as the basic form of instruction. Can research on how children learn basic number facts help in resolving this debate?

A substantial body of research has investigated how children move from their early preference for counting in solving addition and subtraction problems to the adult pattern of retrieving these answers from memory. The contrast of young children's and adults' strategies is based primarily on patterns of reaction times to sets of problems. It is now widely accepted that retrieval rather than counting is the dominant arithmetic strategy for adults and that children who are developing normally move gradually toward the adult strategy between 7 and 11 or 12 years of age. A series of studies by Siegler and Shrager (1984) has shown that children possess a set of strategies that are organized into a hierarchy of preference. Children, like adults, prefer to retrieve answers to arithmetic problems when they can, presumably because of general preference for low effort and quick responses. But children also have individual criteria for certainty and speed of response. If they are unable to retrieve an answer with adequate certainty or speed, they will use one of their backup strategies—mostly involving counting—to produce an answer.

According to Siegler's theory of how this strategy choice works, children have a repertoire of associations of different strengths for different problems. When a problem is presented, they try to retrieve an answer and may

simultaneously gear up for a counting or estimation strategy. If the associative answer is retrieved quickly enough, the other strategies will abort. Each successful answer to a particular problem adds a bit of associative strength. Over time, the associative strength of more and more problems builds to a point where answers are retrieved so quickly that backup strategies are rarely needed. At that point, children resemble adults in their patterns of arithmetic retrieval.

We cannot tell from this research how important school drill on number facts might be in promoting the transition to adultlike patterns of retrieval. No comparable research has been done on unschooled or little-schooled populations of children, and no attempt has been made to control for different school programs. Theoretically, however, there is good reason to believe that suppressing children's informal methods of arithmetic is a poor idea. Counting methods of calculation provide a reliable way for children to generate answers for themselves to any arithmetic problems they may encounter. This is likely to be productive both motivationally (because of the success experienced) and conceptually (because of the number experimentation that is possible). It may also speed up the process of memorization, because there will be more correct answers, and associative strengths may build up more quickly.

Some psychologists have proposed that children should be taught counting-on methods for addition and subtraction early in the school grades (Fuson & Secada, 1986). By teaching counting-on procedures, they argue, not only will addition and subtraction facts be acquired more quickly but so will conceptual understanding. Presumably this would occur because children would spontaneously integrate the counting procedures with their protoquantitative knowledge, just as children who invent counting on for themselves apparently do. Others are skeptical. They are concerned that dependence on counting directs children's attention away from the additive composition properties of numbers. These investigators have proposed alternative early teaching strategies that explicitly direct children's attention to the additive properties of numbers.

For most children, it may not matter very much which approach is used, as long as the counting methods they invent are not actively suppressed. However, for some—those who show mathematics learning difficulties of some kind—the question is more acute. There is some evidence that children who have difficulty learning mathematics are likely to rely on counting methods for a long time. Such reliance would almost certainly interfere with the automatic knowledge of basic addition facts characteristic of children proceeding normally in arithmetic development and might interfere with conceptual development as well. Some children with mathematics disability have been shown to profit substantially from speeded memorization practice, in that they began using strategies more like those of normally progressing children. We do not know the effects of this training on the children's

conceptual knowledge, however. Just as important, no studies yet have examined systematically whether more directly conceptual approaches to teaching basic arithmetic might be even more effective for such children.

Conclusion: directions for the future

The accumulated body of research on the development of children's number knowledge indicates that certain fundamental concepts are normally constructed by children and are, therefore, available as the basis for further mathematical development. Current school practice, however, seems not to build on this informal knowledge, and in some cases, it even suppresses it deliberately. A broad conclusion suggested by this research is that elementary-school teaching focus less on computational drill and more on understanding why arithmetic procedures work even though problem solving would probably be more effective in promoting long-term mathematical achievement among children. Support for this proposal comes from studies comparing Japanese and Chinese elementary-school mathematics instruction with typical American approaches (Stigler & Baranes, in press). These studies have shown that Asian children spend a great deal more time on schoolwork—in school, at home, and in special after-school programs—than do American children, a process supported by important cultural differences. They have also shown major differences between Japanese and American classrooms in the style of mathematics teaching. An elementary school lesson in Japan is likely to consider only two or three problems, discussing them from many angles and exploring underlying principles and implications. A comparable American lesson is likely to spend only a brief time on explaining a procedure and then proceed to have children solve many similar problems, emphasizing accuracy and speed rather than understanding. This difference in instructional style may have much to do with the higher performance of Japanese children in later mathematics achievement that has been repeatedly demonstrated in cross-national comparisons.

Although we can fruitfully use studies of other countries' educational practices as a lens for examining our own, instructional programs cannot be lifted intact out of the cultures that generated them. Instead, we will need to develop and study approaches to mathematics education specifically tuned to our own students and out own cultural needs. For example, in America appropriate forms of practice and conceptual discussion are needed for classes that include children from very diverse family backgrounds. Further, realistic plans must be developed that encourage parental participation but do not expect parents to take over functions, such as supervising arithmetic practice, that Americans have traditionally delegated to the schools.

The specific problems of socially defined groups in America who traditionally do poorly in mathematics and as a result are excluded from certain vocational and career options must also be addressed. Many have proposed

that deficiencies in mathematics performance result primarily from social expectations communicated to children which block motivation for effective mathematics study among girls and members of certain ethnic groups who are not expected to like or do well in mathematics. The cognitive research considered in this article has not, for the most part, been designed specifically to detect gender or ethnic differences in performance. However, children of both sexes and many ethnic groups have been included in the populations studied. The studies have not typically found differences between boys and girls or between social class and ethnic groups in the kinds of mathematical concepts that are developed, although socially less privileged children may develop them a bit more slowly.

There is good reason to believe—although no hard evidence to prove—that children from families most alienated from schooling and school culture are even less likely than those from the mainstream culture to trust their own intuitive and informal mathematical ideas inside the school setting. The result is that they, even more than others, treat school mathematics as a matter of learning symbol-manipulation procedures, with the kind of negative results discussed above. Girls, too, may be more apt to do this than boys, probably because of cultural expectations for them. If this were so, a general approach to school mathematics instruction that stressed concepts and explicitly engaged children's informally developed knowledge might be expected to yield particular benefits for minority children and perhaps for girls as well. Cognitive research, in other words, points to the need for a general reorientation of early mathematics instruction, and there is reason to believe that such a reorientation would benefit all children.

References

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology*, *3*, 21–29.

Fuson, K. C., & Secada, W. G. (1986). Teaching children to add by counting-on with one-handed finger patterns. *Cognition and Instruction*, *3*, 229–260.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gelman, R., & Greeno, J. G. (in press). On the nature of competence: Principles for understanding in a domain. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation*. New York: Norton.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*, 109–129.

Nesher, P. (1988). Multiplicative school word problems: Theoretical approaches and empirical findings. In M. J. Behr & J. Hiebert (Eds.), *Research agenda in mathematical education: Number concepts and operations in the middle grades*. Reston, VA: National Council of Teachers of Mathematics.

Piaget, J. (1965). *The child's conception of number*. New York: Norton. (Original work published 1941)

Resnick, L. B. (1983). A developmental theory of number understanding. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 109–151). New York: Academic Press.

Resnick, L. B. (1986). The development of mathematical intuition. In M. Perlmutter (Ed.), *Perspectives on intellectual development: The Minnesota Symposia on Child Psychology* (Vol. 19, pp. 159–194). Hillsdale, NJ: Erlbaum.

Resnick, L. B. (1987). Constructing knowledge in school. In L. S. Liben (Ed.), *Development and learning: Conflict or congruence?* (pp. 19–50). Hillsdale, NJ: Erlbaum.

Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and solving problems. *Cognition and Instruction*, *5*, 49–101.

Saxe, G. B., Guberman, S. R., & Gearhart, M. (1987). Social processes in early number development. *Monographs of the Society for Research in Child Development*, *52*(2, Serial No. 216).

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, *46*(2, Serial No. 189).

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

Sophian, C. (1987). Early developments in children's use of counting to solve quantitative problems. *Cognition and Instruction*, *4*, 61–90.

Starkey, P., Spelke, E. S., & Gelman, R. (in press). Numerical abstraction by human infants. *Cognition*.

Stigler, J. W., & Baranes, R. (in press). Culture and mathematics learning. In E. Rothkopf (Ed.), *Review of research in education*. Washington, DC: American Educational Research Association.

70

MATHEMATICS IN THE STREETS AND IN SCHOOLS

T. Nunes Carraher, D. W. Carraher and A. D. Schliemann

An analysis of everyday use of mathematics by working youngsters in commercial transactions in Recife, Brazil, revealed computational strategies different from those taught in schools. Performance on mathematical problems embedded in real-life contexts was superior to that on school-type word problems and context-free computational problems involving the same numbers and operations. Implications for education are examined.

There are reasons for thinking that there may be a difference between solving mathematical problems using algorithms learned in school and solving them in familiar contexts out of school. Reed & Lave (1981) have shown that people who have not been to school often solve such problems in different ways from people who have. This certainly suggests that there are informal ways of doing mathematical calculations which have little to do with the procedures taught in school.

Reed & Lave's study with Liberian adults showed differences between people who had and who had not been to school. However, it is quite possible that the same differences between informal and school-based routines could exist within people. In other words it might be the case that the same person could solve problems sometimes in formal and at other times in informal ways. This seems particularly likely with children who often have to do mathematical calculations in informal circumstances outside school at the same time as their knowledge of the algorithms which they have to learn at school is imperfect and their use of them ineffective.

We already know that children often obtain absurd results such as finding a remainder which is larger than the minuend when they try to apply routines for computations which they learn at school (Carraher & Schliemann, in press). There is also some evidence that informal procedures learned outside

Source: *British Journal of Developmental Psychology*, 1985, 3(1), 21–29.

school are often extremely effective. Gay & Cole (1976) for example showed that unschooled Kpelle traders estimated quantities of rice far better than educated Americans managed to. So it seems quite possible that children might have difficulty with routines learned at school and yet at the same time be able to solve the mathematical problems for which these routines were devised in other more effective ways. One way to test this idea is to look at children who have to make frequent and quite complex calculations outside school. The children who sell things in street markets in Brazil form one such group (Carraher *et al*., 1982).

The cultural context

The study was conducted in Recife, a city of approximately 1.5 million people on the north-eastern coast of Brazil. Like several other large Brazilian cities, Recife receives a very large number of migrant workers from the rural areas who must adapt to a new way of living in a metropolitan region. In an anthropological study of migrant workers in São Paulo, Brazil, Berlinck (1977) identified four pressing needs in this adaptation process: finding a home, acquiring work papers, getting a job, and providing for immediate survival (whereas in rural areas the family often obtains food through its own work). During the initial adaptation phase, survival depends mostly upon resources brought by the migrants or received through begging. A large portion of migrants later become unspecialized manual workers, either maintaining a regular job or working in what is known as the informal sector of the economy (Cavalcanti, 1978). The informal sector can be characterized as an unofficial part of the economy which consists of relatively unskilled jobs not regulated by government organs thereby producing income not susceptible to taxation while at the same time not affording job security or workers' rights such as health insurance. The income generated thereby is thus intermittent and variable. The dimensions of a business enterprise in the informal sector are determined by the family's work capability. Low educational and professional qualification levels are characteristic of the rather sizable population which depends upon the informal sector. In Recife, approximately 30 per cent of the workforce is engaged in the informal sector as its main activity and 18 per cent as a secondary activity (Cavalcanti, 1978). The importance of such sources of income for families in Brazil's lower socio-economic strata can be easily understood by noting that the income of an unspecialized labourer's family is increased by 56 per cent through his wife's and children's activities in the informal sector in São Paulo (Berlinck, 1977). In Fortaleza it represents fully 60 per cent of the lower class[1] family's income (Cavalcanti & Duarte, 1980*a*).

Several types of occupations—domestic work, street-vending, shoe-repairing and other types of small repairs which are carried out without a fixed commercial address—are grouped as part of the informal sector of the economy.

The occupation considered in the present study—that of street-vendors—represents the principal occupation of 10 per cent of the economically active population of Salvador (Cavalcanti & Duarte, 1980*b*) and Fortaleza (Cavalcanti & Duarte, 1980*a*). Although no specific data regarding street-vendors were obtained for Recife, data from Salvador and Fortaleza serve as close approximations since these cities are, like Recife, State capitals from the same geographical region.

It is fairly common in Brazil for sons and daughters of street-vendors to help out their parents in their businesses. From about the age of 8 or 9 the children will often enact some of the transactions for the parents when they are busy with another customer or away on some errand. Pre-adolescents and teenagers may even develop their own 'business', selling snack foods such as roasted peanuts, pop-corn, coconut milk or corn on the cob. In Fortaleza and Salvador, where data are available, 2.2 and 1.4 per cent, respectively, of the population actively engaged in the informal sector as street-vendors were aged 14 or less while 8.2 and 7.5 per cent, respectively, were aged 15–19 years (Cavalcanti & Duarte, 1980*a*,*b*).

In their work these children and adolescents have to solve a large number of mathematical problems, usually without recourse to paper and pencil. Problems may involve multiplication (one coconut cost x; four coconuts, $4x$), addition (4 coconuts and 12 lemons cost $x + y$), and subtraction (Cr$ 500—i.e. 500 *cruzeiros*—minus the purchase price will give the change due). Division is much less frequently used but appears in some contexts in which the price is set with respect to a measuring unit (such as 1 kg) and the customer wants a fraction of that unit: for example, when the particular item chosen weighs 1.2 kg. The use of tables listing prices by number of items (one egg—12 *cruzeiros*; two eggs—24, etc.) is observed occasionally in natural settings but was not observed among the children who took part in the study. Pencil and paper were also not used by these children, although they may occasionally be used by adult vendors when adding long lists of items.

Method

Subjects

The children in this study were four boys and one girl aged 9–15 years with a mean age of 11.2 and ranging in level of schooling from first to eighth grade. One of them had only one year of schooling; two had three years of schooling; one, four years; and one, eight years. All were from very poor backgrounds. Four of the subjects were attending school at the time and one had been out of school for two years. Four of these subjects had received formal instruction on mathematical operations and word problems. The subject who attended first grade and dropped out of school was unlikely to have learned multiplication and division in school since these operations are usually initiated in second or third grade in public schools in Recife.

Procedure

The children were found by the interviewers on street corners or at markets where they worked alone or with their families. Interviewers chose subjects who seemed to be in the desired age range—school children or young adolescents—obtaining information about their age and level of schooling along with information on the prices of their merchandise. Test items in this situation were presented in the course of a normal sales transaction in which the researcher posed as a customer. Purchases were sometimes carried out. In other cases the 'customer' asked the vendor to perform calculations on possible purchases. At the end of the informal test, the children were asked to take part in a formal test which was given on a separate occasion, no more than a week later, by the same interviewer. Subjects answered a total of 99 questions on the formal test and 63 questions on the informal test. Since the items of the formal test were based upon questions of the informal test, order of testing was fixed for all subjects.

(1) The informal test

The informal test was carried out in Portuguese in the subject's natural working situation, that is, at street corners or an open market. Testers posed to the subject successive questions about potential or actual purchases and obtained verbal responses. Responses were either tape-recorded or written down, along with comments, by an observer. After obtaining an answer for the item, testers questioned the subject about his or her method for solving the problem.

The method can be described as a hybrid between the Piagetian clinical method and participant observation. The interviewer was not merely an interviewer; he was also a customer—a questioning customer who wanted the vendor to tell him how he or she performed their computations.

An example is presented below taken from the informal test with M., a coconut vendor aged 12, third grader, where the interviewer is referred to as 'customer':

Customer: How much is one coconut?
M: 35.
Customer: I'd like ten. How much is that?
M: (Pause) Three will be 105; with three more, that will be 210. (Pause) I need four more. That is . . .[2] (pause) 315 . . . I think it is 350.

This problem can be mathematically represented in several ways: 35×10 is a good representation of the *question* posed by the interviewer. The subject's answer is better represented by $105 + 105 + 105 + 35$, which implies that 35×10 was solved by the subject as $(3 \times 35) + (3 \times 35) + (3 \times 35) + 35$. The subject can be said to have solved the following subitems in the above situation:

(*a*) 35×10;
(*b*) 35×3 (which may have already been known);
(*c*) $105 + 105$;
(*d*) $210 + 105$;
(*e*) $315 + 35$;
(*f*) $3 + 3 + 3 + 1$.

When one represents in a formal mathematical fashion the problems which were solved by the subject, one is in fact attempting to represent the subject's mathematical competence. M. proved to be competent in finding out how much 35×10 is, even though he used a routine not taught in third grade, since in Brazil third-graders learn to multiply any number by ten simply by placing a zero to the right of that number. Thus, we considered that the subject solved the test item (35×10) and a whole series of sub-items (*b* to *f*) successfully in this process. However, in the process of scoring, only *one* test item (35×10) was considered as having been presented and, therefore, correctly solved.

(2) The formal test

After subjects were interviewed in the natural situation, they were asked to participate in the formal part of the study and a second interview was scheduled at the same place or at the subject's house.

The items for the formal test were prepared for each subject on the basis of problems solved by him or her during the informal test. Each problem solved in the informal test was mathematically represented according to the subject's problem-solving routine.

From all the mathematical problems *successfully solved* by each subject (regardless of whether they constituted a test item or not), a sample was chosen for inclusion in the subject's formal test. This sample was presented in the formal test either as a mathematical operation dictated to the subject (e.g. 105 + 105) or as a word problem e.g. Mary bought *x* bananas; each banana cost *y*; how much did she pay altogether?). In either case, *each subject solved problems employing the same numbers involved in his or her own informal test*. Thus quantities used varied from one subject to the other.

Two variations were introduced in the formal test, according to methodological suggestions contained in Reed & Lave (1981). First, some of the items presented in the formal test were the inverse of problems solved in the informal test (e.g. 500–385 may be presented as 385 + 115 in the formal test). Second, some of the items in the informal test used a decimal value which differed from the one used in the formal test (e.g. 40 *cruzeiros* may have appeared as 40 *centavos* or 35 may have been presented as 3500 in the formal test—the principal Brazilian unit of currency is the *cruzeiro*; each *cruzeiro* is worth one hundred *centavos*).

In order to make the formal test situation more similar to the school setting, subjects were given paper and pencil at the testing and were encouraged to use these. When problems were nonetheless solved without recourse to writing, subjects were asked to write down their answers. Only one subject refused to do so, claiming that he did not know how to write. It will be recalled, however, that the school-type situation was not represented solely by the introduction of pencil and paper but also by the very use of formal mathematical problems without context and by word problems referring to imaginary situations.

In the formal test the children were given a total of 38 mathematical operations and 61 word problems. Word problems were rather concrete and each involved only one mathematical operation.

Results and discussion

The analysis of the results from the informal test required an initial definition of what would be considered a test item in that situation. While, in the formal test, items were defined prior to testing, in the informal test problems were generated in the natural setting and items were identified *a posteriori*. In order to avoid a biased increase in the number of items solved in the informal test, the definition of an item was based upon *questions* posed by the customer/tester. This probably constitutes a conservative estimate of the number of problems solved, since subjects often solved a number of intermediary steps in the course of searching for the solution to the question they had been asked. Thus the same defining criterion was applied in both testing situations in the identification of items even though items were defined prior to testing in one case and after testing in the other. In both testing situations, the subject's oral response was the one taken into account even though in the formal test written responses were also available.

Context-embedded problems were much more easily solved than ones without a context. Table 1 shows that 98.2 per cent of the 63 problems presented in the informal test were correctly solved. In the formal test word problems (which provide some descriptive context for the subject), the rate of correct responses was 73.7 per cent, which should be contrasted with a 36.8 per cent rate of correct responses for mathematical operations with no context.

The frequency of correct answers for each subject was converted to scores from 1 to 10 reflecting the percentage of correct responses. A Friedman two-way analysis of variance of score ranks compared the scores of each subject in the three types of testing conditions. The scores differ significantly across conditions ($\chi^2 r = 6.4$, $P = 0.039$). Mann–Whitney *Us* were also calculated comparing the three types of testing situations. Subjects performed significantly better on the informal test than on the formal test involving context-free operations ($U = 0$, $P < 0.05$). The difference between the informal test and the word problems was not significant ($U = 6$, $P > 0.05$).

Table 1 Results according to testing conditions

	Informal test		*Formal test*			
			Mathematical operations		*Word problems*	
Subject	*Score*[a]	*Number of items*	*Score*	*No. items*	*Score*	*No. items*
M	10	18	2.5	8	10	11
P	8.9	19	3.7	8	6.9	16
Pi	10	12	5.0	6	10	11
MD	10	7	1.0	10	3.3	12
S	10	7	8.3	6	7.3	11
Totals		63		38		61

[a] Each subject's score is the percentage of correct items divided by 10.

It could be argued that errors observed in the formal test were related to the transformations that had been performed upon the informal test problems in order to construct the formal test. An evaluation of this hypothesis was obtained by separating items which had been changed either by inverting the operation or changing the decimal point from items which remained identical to their informal test equivalents. The percentage of correct responses in these two groups of items did not differ significantly; the rate of correct responses in transformed items was slightly higher than that obtained for items identical to informal test items. Thus the transformations performed upon informal test items in designing formal test items cannot explain the discrepancy of performance in these situations.

A second possible interpretation of these results is that the children interviewed in this study were 'concrete' in their thinking and, thus, concrete situations would help them in the discovery of a solution. In the natural situation, they solved problems about the sale of lemons, coconuts, etc., when the actual items in question were physically present. However, the presence of concrete instances can be understood as a facilitating factor if the instance somehow allows the problem-solver to abstract from the concrete example to a more general situation. There is nothing in the nature of coconuts that makes it relatively easier to discover that three coconuts (at Cr$ 35.00 each) cost Cr$ 105.00. The presence of the groceries does not simplify the arithmetic of the problem. Moreover, computation in the natural situation of the informal test was in all cases carried out mentally, without recourse to external memory aids for partial results or intermediary steps. One can hardly argue that mental computation would be an ability characteristic of concrete thinkers.

The results seem to be in conflict with the implicit pedagogical assumption of mathematical educators according to which children ought first to learn mathematical operations and only later to apply them to verbal and real-life problems. Real-life and word problems may provide the 'daily human sense' (Donaldson, 1978) which will guide children to find a correct solution intuitively without requiring an extra step—namely, the translation of word problems into algebraic expressions. This interpretation is consistent with data obtained by others in the area of logic, such as Wason & Shapiro (1971), Johnson-Laird *et al.* (1972) and Lunzer *et al.* (1972).

How is it possible that children capable of solving a computational problem in the natural situation will fail to solve the same problem when it is taken out of its context? In the present case, a qualitative analysis of the protocols suggested that the problem-solving routines used may have been different in the two situations. In the natural situations children tended to reason by using what can be termed a 'convenient group' while in the formal test school-taught routines were more frequently, although not exclusively, observed. Five examples are given below, which demonstrate the children's ability to deal with quantities and their lack of expertise in manipulating symbols. The examples were chosen for representing clear explanations of the procedures used in both settings. In each of the five examples below the performance described in the informal test contrasts strongly with the same child's performance in the formal test when solving the same item.

(1) First example (M, 12 years)

Informal test

Customer: I'm going to take four coconuts. How much is that?
Child: Three will be 105, plus 30, that's 135 . . . one coconut is 35 . . . that is . . . 140!

Formal test

Child resolves the item 35×4 explaining out loud:
4 times 5 is 20, carry the 2; 2 plus 3 is 5, times 4 is 20.
Answer written: 200.

(2) Second example (MD, 9 years)

Informal test

Customer: OK, I'll take three coconuts (at the price of Cr$ 40.00 each). How much is that?
Child: (Without gestures, calculates out loud) 40, 80, 120.

Formal test

Child solves the item 40×3 and obtains 70. She then explains the procedure 'Lower the zero; 4 and 3 is 7'.

(3) Third example (MD, 9 years)

Informal test
Customer: I'll take 12 lemons (one lemon is Cr$ 5.00).
Child: 10, 20, 30, 40, 50, 60 (while separating out two lemons at a time).

Formal test
Child has just solved the item 40×3. In solving 12×5 she proceeds by lowering first the 2, then the 5 and the 1, obtaining 152. She explains this procedure to the (surprised) examiner when she is finished.

(4) Fourth example (S, 11 years)

Informal test
Customer: What would I have to pay for six kilos? (of watermelon at Cr$ 50.00 per kg).
Child: [Without any appreciable pause] 300.
Customer: Let me see. How did you get that so fast?
Child: Counting one by one. Two kilos, 100. 200. 300.

Formal test
Test item: A fisherman caught 50 fish. The second one caught five times the amount of fish the first fisherman had caught. How many fish did the lucky fisherman catch?
Child: (Writes down 50×6 and 360 as the result; then answers) 36.
Examiner repeats the problems and child does the computation again, writing down 860 as result. His oral response is 86.
Examiner: How did you calculate that?
Child: I did it like this. Six times six is 36. Then I put it there.
Examiner: Where did you put it? (Child had not written down the number to be carried.)
Child: (Points to the digit 5 in 50). That makes 86 [apparently adding 3 and 5 and placing this sum in the result].
Examiner: How many did the first fisherman catch?
Child: 50.

A final example follows, with suggested interpretations enclosed in parentheses.

(5) Fifth example

Informal test
Customer: I'll take two coconuts (at Cr$ 40.00 each. Pays with a Cr$ 500.00 bill). What do I get back?
Child: (Before reaching for customer's change) 80, 90, 100. 420.

Formal test
Test item: 420 + 80.
The child writes 420 plus 80 and claims that 130 is the result. [The procedure used was not explained but it seems that the child applied a step of a multiplication routine to an addition problem by successively adding 8 to 2 and then to 4, carrying the 1; that is, 8 + 2 = 10, carry the one, 1 + 4 + 8 = 13. The zeros in 420 and 80 were not written. Reaction times were obtained from tape recordings and the whole process took 53 seconds.]
Examiner: How did you do this one, 420 plus 80?
Child: Plus?
Examiner: Plus 80.
Child: 100, 200.
Examiner: (After a 5 second pause, interrupts the child's response treating it as final) Hum, OK.
Child: Wait a minute. That was wrong. 500. [The child had apparently added 80 and 20, obtaining one hundred, and then started adding the hundreds. The experimenter interpreted 200 as the final answer after a brief pause but the child completed the computation and gave the correct answer when solving the addition problem by a manipulation-with-quantities approach.]

In the informal test, children rely upon mental calculations which are closely linked to the quantities that are being dealt with. The preferred strategy for multiplication problems seems to consist in chaining successive additions. In the first example, as the addition became more difficult, the subject decomposed a quantity into tens and units—to add 35 to 105, M. first added 30 and later included 5 in the result.

In the formal test, where paper and pencil were used in all the above examples, the children try to follow, without success, school-prescribed routines. Mistakes often occur as a result of confusing addition routines with multiplication routines, as is clearly the case in examples (1) and (5). Moreover, in all the cases, there is no evidence, once the numbers are written down, that the children try to relate the obtained results to the problem at hand in order to assess the adequacy of their answers.

Summarizing briefly, the combination of the clinical method of questioning with participant observation used in this project seemed particularly helpful when exploring mathematical thinking and thinking in daily life. The results support the thesis proposed by Luria (1976) and by Donaldson (1978) that thinking sustained by daily human sense can be—in the same subject—at a higher level than thinking out of context. They also raise doubts about the pedagogical practice of teaching mathematical operations in a disembedded form before applying them to word problems.

Our results are also in agreement with data reported by Lave *et al*. (1984), who showed that problem solving in the supermarket was significantly

superior to problem solving with paper and pencil. It appears that daily problem solving may be accomplished by routines different from those taught in schools. In the present study, daily problem solving tended to be accomplished by strategies involving the mental manipulation of quantities while in the school-type situation the manipulation of symbols carried the burden of computation, thereby making the operations 'in a very real sense divorced from reality' (see Reed & Lave, 1981, p. 442). In many cases attempts to follow school-prescribed routines seemed in fact to interfere with problem solving (see also Carraher & Schliemann, in press).

Are we to conclude that schools ought to allow children simply to develop their own computational routines without trying to impose the conventional systems developed in the culture? We do not believe that our results lead to this conclusion. Mental computation has limitations which can be overcome through written computation. One is the inherent limitation placed on multiplying through successive chunking, that is, on multiplying through repeated chunked additions—a procedure which becomes grossly inefficient when large numbers are involved.

The sort of mathematics taught in schools has the potential to serve as an 'amplifier of thought processes', in the sense in which Bruner (1972) has referred to both mathematics and logic. As such, we do not dispute whether 'school maths' routines can offer richer and more powerful alternatives to maths routines which emerge in non-school settings. The major question appears to centre on the proper pedagogical point of departure, i.e. where to start. We suggest that educators should question the practice of treating mathematical systems as formal subjects from the outset and should instead seek ways of introducing these systems in contexts which allow them to be sustained by human daily sense.

Acknowledgements

The research conducted received support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasília, and from the British Council.

The authors thank Peter Bryant for his helpful comments on the present report.

Notes

1 In the present report the term 'class' is employed loosely, without a clear distinction from the expression 'socio-economic stratum'.

2 (. . .) is used here to mark ascending intonation suggestive of the interruption, and not completion, of a statement.

References

Berlinck, M. T. (1977). *Marginalidade Social e Relações de Classe em São Paulo*. Petrópolis, RJ, Brazil: Vozes.

Bruner, J. (1972). *Relevance of Education*. London: Penguin.
Carraher, T., Carraher, D. & Schliemann, A. (1982). Na vida dez, na escola zero: Os contextos culturais da aprendizagem da matemática. *Cadernos de Pesquisa*, **42**, 79–86. (São Paulo, Brazil, special UNESCO issue for Latin America.)
Carraher, T. & Schliemann, A. (in press). Computation routines prescribed by schools: Help or hindrance? *Journal for Research in Mathematics Education*.
Cavalcanti, C. (1978). *Viabilidade do Setor Informal. A Demanda de Pequenos Serviços no Grande Recife*. Recife, PE, Brazil: Instituto Joaquim Nabuco de Pesquisas Sociais.
Cavalcanti, C. & Duarte, R. (1980*a*). *A Procura de Espaço na Economia Urbana: O Setor Informal de Fortaleza*. Recife, PE, Brazil: SUDENE/FUNDAJ.
Cavalcanti, C. & Duarte, R. (1980*b*). *O Setor Informal de Salvador: Dimensões, Natureza, Significação*. Recife, PE, Brazil: SUDENE/FUNDAJ.
Donaldson, M. (1978). *Children's Minds*. New York: Norton.
Gay, J. & Cole, M. (1976). *The New Mathematics and an Old Culture: A Study of Learning among the Kpelle of Liberia*. New York: Holt, Rinehart & Winston.
Johnson-Laird, P. N., Legrenzi, P. & Sonino Legrenzi, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, **63**, 395–400.
Lave, J., Murtaugh, M. & de La Rocha, O. (1984). The dialectical construction of arithmetic practice. In B. Rogoff & J. Lave (eds), *Everyday Cognition: Its Development in Social Context*, pp. 67–94. Cambridge, MA: Harvard University Press.
Lunzer, E. A., Harrison, C. & Davey, M. (1972). The four-card problem and the development of formal reasoning. *Quarterly Journal of Experimental Psychology*, **24**, 326–339.
Luria, A. R. (1976). *Cognitive Development: Its Cultural and Social Foundations*. Cambridge, MA: Harvard University Press.
Reed, H. J. & Lave, J. (1981). Arithmetic as a tool for investigating relations between culture and cognition. In R. W. Casson (ed.), *Language, Culture and Cognition: Anthropological Perspectives*. New York: Macmillan.
Wason, P. C. & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, **23**, 63–71.

71

FOSTERING COGNITIVE GROWTH

A perspective from research on mathematics learning and instruction

E. De Corte

There is now robust empirical evidence that shows that learning and teaching in schools make a substantial difference in students' cognitive growth. Taking this as a starting point and focusing on mathematics as a subject-matter domain, this article addresses the crucial issue of elaborating a coherent framework for the design of learning environments that can elicit and maintain in all students the acquisition processes that are conducive to the intended cognitive growth. On the basis of recent research on mathematics learning and instruction, I argue that the design of such environments should be guided by (a) the conception that the ultimate objective of mathematics education is the acquisition of a mathematical disposition and (b) a constructivist view of mathematics learning as the interactive, cumulative, and situated construction of knowledge, skills, beliefs and attitudes mediated by the teacher. Design principles for powerful learning environments that derive from these perspectives on mathematics education are illustrated by a brief description of the major characteristics of one innovative project for mathematics teaching at the primary school: Realistic Mathematics Education.

As Weinert and Helmke (1995) note in their article, Ceci (1991) concluded that schools do make a difference with respect to cognitive development, in the sense that the acquisition and growth of the cognitive skills and processes underlying intellectual performance are, to a large degree, the result of learning and teaching in school.

Similar findings were reported by Husen and Tuijnman (1991), based on a LISREL reanalysis of a large data set of a longitudinal Swedish study. They concluded that formal schooling plays a crucial role in enhancing the

Source: *Educational Psychologist*, 1995, 30(1), 37–46.

intellectual capital of a nation and therefore in increasing the number of youngsters who can profit from further education.

These research outcomes strongly contradict the claim derived from older, influential work (e.g., Coleman et al., 1966; Jenks et al., 1972), that schools do not have a substantial impact on the development of educational abilities. At the same time, these outcomes have implications for educational policy, instructional research, and classroom practice. More specifically, with respect to research on learning and instruction, these findings confront us with a challenging task: the elaboration of a framework consisting of coherent, research-based principles for the design of powerful learning environments (i.e., situations and contexts that can elicit in all students learning and developmental processes that result in an increase of their cognitive potential).

Two major aspects of such environments concern what should be taught and learned, and how it should be taught and learned. Both aspects are discussed in this article, focusing thereby on the domain of learning and teaching mathematics problem solving to elementary school children.

Acquiring a mathematical disposition

The analysis of problem-solving expertise in a large variety of domains, including mathematics, has led to the identification and more precise definition of the crucial aptitudes involved in competent learning and problem solving. The term *aptitude* is used here in a broad sense: It refers to any characteristic of a student that can influence his or her learning and problem-solving activity and achievement (see Snow, 1992). With respect to mathematics, there is nowadays a rather broad consensus that the major categories of aptitudes underlying skilled problem solving are domain-specific knowledge, heuristic methods, metacognitive knowledge and skills, and affective components, especially beliefs and emotions (see Schoenfeld, 1992). Thus, good performance in mathematics requires more than the acquisition of the variety of procedural computational skills that have prevailed—and often still prevail—in mathematics teaching (for a more detailed discussion, see De Corte, Greer, & Verschaffel, in press).

Domain-specific knowledge

Domain-specific knowledge involves facts, symbols, conventions, definitions, formulas, algorithms, concepts, and rules, which constitute the substance or the content of a subject-matter field. A major finding of the analysis of expertise is that expert problem solvers master a large, well-organized, and flexibly accessible domain-specific knowledge base (Chi, Glaser, & Farr, 1988).

But conceptual domain-specific knowledge already strongly affects the solution processes of young children on one-step addition and subtraction word

Table 1 Results of 30 first graders on three addition word problems at the beginning of the school year

Problem	*Problem structure*	*Number of correct solutions*
1. Pete has 3 apples; Ann has 7 apples. How many apples do they have all together?	Combine: whole set unknown	26
2. Pete has some apples. He gave 3 apples to Ann; now Pete has 5 apples. How many apples did Pete have in the beginning?	Change: initial set unknown	12
3. Pete has 3 apples; Ann has 6 more apples than Pete. How many apples does Ann have?	Compare: compared set unknown	5

problems. For example, as illustrated in Table 1, De Corte and Verschaffel (1987) found substantial differences in difficulty level between word problems that could be solved by the same arithmetic operation but represented different categories of problem situations (see Fuson, 1992). These findings show that, to understand and solve even those simple word problems, it is not sufficient to master the arithmetic operations of addition and subtraction; children must also apply conceptual knowledge of the underlying problem structures.

The importance of domain-specific knowledge is also convincingly supported in a negative way by many research findings that show the occurrence of misconceptions and defective skills in many learners. For instance, the so-called multiplication-makes-bigger misconception has been observed in students of different ages and in a diversity of countries (see De Corte, Verschaffel, & Van Coillie, 1988; Greer, 1992).

Even more crucial than mastering separate pieces of subject-matter content is the availability and accessibility of a well-organized knowledge base. Indeed, it has been shown that experts differ from novices in that their knowledge base is better and more dynamically structured, and as a consequence more flexibly accessible (Chi et al., 1988).

Heuristic methods

Heuristic methods are systematic search strategies for problem analysis and transformation. They do not guarantee that one will find the solution of a given problem; however, because they induce a systematic and planned approach to the task—in contrast to a trial-and-error strategy—heuristic methods substantially increase one's probability of success in solving the problem. Some examples of heuristic methods are carefully analyzing a problem by specifying the knowns and the unknowns, decomposing the problem into

subgoals, finding an easier related or analogous problem, visualizing the problem with a drawing or a diagram, working backward from the intended goal or solution, and provisionally relaxing one of the constraints of the solution and returning later to reimpose it.

One major way in which heuristics can be helpful in solving a problem is as tools or resources that the problem solver uses in transforming the original problem so that a familiar, routine task emerges for which he or she has a ready-made solution. Consider the following example: "A store sells two kinds of fruit juice: Bottle A costs 16 Belgian francs for 20 centiliters, and Bottle B is priced 19 francs for 25 centiliters. What is the best buy, assuming that both kinds of juice are of equal quality?" In solving this problem, a student might think of a related task solved before, such as comparing the price of potatoes in sacks of different weights. Through the analogy of figuring out the price per kilogram of each kind of potatoes, the student might decide also to decompose the present problem by calculating first the price per liter for each type of fruit juice, and then comparing both prices, which is of course a routine task.

Metacognitive knowledge and skill

Metacognition involves two main aspects: knowledge concerning one's own cognitive functioning and activities relating to the self-monitoring of one's cognitive processes (Brown, Bransford, Ferrara, & Campione, 1983).

Metacognitive knowledge includes knowing about the strengths as well as the weaknesses and limits of one's cognitive capacities; for example, being aware of the limits of short-term memory and knowing that our memory is fallible but that one can use aids (e.g., mnemonics) for retaining certain information. Beliefs about cognition and ability are also involved. The findings of Dweck and Elliott (1983) are important in this respect and with regard to learning in general. According to Dweck and Elliott, the specific actions that individuals take in a learning or problem-solving situation depend on the particular conception about ability that they hold. They found two very different conceptions of ability or theories of intelligence in children. The *entity* conception considers ability to be a global, stable, and unchangeable characteristic reflected in one's performance, whereas the *incremental* conception treats ability as a set of skills that can be expanded and improved through learning and effort. It is obvious that both groups of children have different motivations for, and approaches to, new learning tasks and problems.

The self-monitoring or self-regulation mechanisms that constitute the second component of metacognition can be defined as the executive control structure that organizes and guides our learning and thinking processes. This includes skills such as planning a solution process, monitoring an ongoing solution process, evaluating and, if necessary, debugging an answer or a solution, and reflecting on one's learning and problem-solving activities.

Evidence to support the crucial role of metacognition in learning and problem solving has been obtained in comparative studies of skilled and weak problem solvers of different ages and in a variety of content domains, including mathematics. For example, Nelissen (1987) found significant differences between high-ability and low-ability elementary and secondary school pupils for self-monitoring and self-control while solving mathematical tasks. Similar outcomes were reported by Overtoom (1991) who compared gifted and average students at the primary and secondary school level.

Affective components

Although it has been recognized for some time that affective factors play an important role in mathematics teaching and learning, the scholarly community failed for a long time to include these factors in its research projects. Recent work, however, has begun to counteract this tendency (Boekaerts, 1993; McLeod, 1990; McLeod & Adams, 1989).

The affective domain involves beliefs, attitudes, and emotions that reflect the whole range of affective reactions involved in mathematics learning (McLeod, 1990). These terms refer to responses that vary in the intensity of affect involved, namely from rather cold for beliefs to hot for emotions. They also differ in terms of stability. Whereas beliefs and attitudes are rather stable and resistant to change, emotions change quickly. Finally, the three affective aspects are distinct with regard to their degree of cognitive loading. Beliefs have a very strong cognitive component that decreases over attitudes toward emotions. It becomes at the same time clear that, although cognition and affect are interwoven in all reactions, beliefs are the most obvious interface. This is also illustrated by the fact that some authors consider beliefs as an aspect of metacognition (e.g., Schoenfeld, 1987).

Research has already identified students' beliefs about mathematics, many of which are induced by instruction and have a negative or inhibitory influence on students' learning activities and approach to mathematics problems (Greeno, 1991b). For example, Schoenfeld (1988) reported that in high-school classes in which mathematics is taught in a way that would generally be considered good teaching, students nevertheless acquire beliefs about the domain, such as, "Solving a math problem should not take more than just a few minutes" or "Being able to solve a math problem is a mere question of luck." It is obvious that such misconceptions will not promote a mindful and persistent approach to new and challenging problems.

On the other hand, a longitudinal study by Helmke (1992) showed that, at the beginning of their school career, children have a very positive attitude toward mathematics. But Helmke also observed a downward trend throughout the primary school, which was stronger in girls than in boys. It is thus important to design mathematics learning environments in such a way that childrens' positive initial attitudes and beliefs do not fade, but are maintained

and stimulated, especially because it is well known that negative attitudes and beliefs are resistant to change.

Interaction among the different categories of aptitudes: toward a dispositional view of mathematics learning

So far, the different aptitudes involved in problem solving have been discussed separately. However, it is obvious that, in expert problem solving, those aptitudes—and also their subcategories—are applied integratively and interactively.

Even solving the one-step addition and subtraction word problems from Table 1 requires mastery of two subcategories of domain-specific knowledge: procedural skills for counting and computing and conceptual knowledge of the underlying structure of the different problem situations.

But the relations among the different categories of aptitudes are also very prominent in skilled problem solving. For example, discovering the applicability of a heuristic such as finding an easier related or analogous problem is usually based on one's conceptual, domain-specific knowledge of the content or topic involved in the problem.

This integration of the different categories of aptitudes is certainly necessary, but still not enough to overcome the well-known phenomenon of inert knowledge observed in many students: Although the knowledge is available and can even be recalled on request, students do not spontaneously apply it in situations where it is relevant to solve new problems. Expertise in mathematics problem solving indeed involves more than the mere sum of the four categories of aptitudes mentioned earlier. In this respect, the notion of a *mathematical disposition* is useful to refer to the integrated availability and application of those aptitudes, as described by the National Council of Teachers of Mathematics (NCTM; 1989):

> Learning mathematics extends beyond learning concepts, procedures, and their applications. It also includes developing a disposition toward mathematics and seeing mathematics as a powerful way for looking at situations. Disposition refers not simply to attitudes but to a tendency to think and to act in positive ways. Students' mathematical dispositions are manifested in the way they approach tasks—whether with confidence, willingness to explore alternatives, perseverance, and interest—and in their tendency to reflect on their own thinking. (p. 233)

This view of expertise in mathematics is in accordance with recent ideas in the more general literature on learning and instruction, such as the dispositional approach to thinking and creativity proposed by Perkins, Jay, and Tishman (1993). According to these authors, the notion of disposition involves more

than ability and motivation, although both are important aspects of it. They distinguish three components of a disposition: inclination, sensitivity, and ability. *Inclination* is the tendency to engage in a given behavior due to motivation, habits, and possibly other factors. *Sensitivity* refers to the feeling for and alertness to opportunities for implementing the appropriate behavior. *Ability*, then, constitutes the actual skill in deploying the behavior. This dispositional conception provides an explanation for the phenomenon of inert knowledge: Students often have the ability to perform certain tasks or solve certain problems, but do not exercise them because of lack of spontaneous inclination and sensitivity.

But how do such aspects as sensitivity and inclination relate to the different kinds of aptitudes? Should they not be considered as an additional category of aptitudes that also has to be pursued as a direct goal of instruction? Although continued research on this issue is needed, the most plausible perspective seems to be that ability, sensitivity, and inclination are characteristics of the aptitudes described before. This implies that it is not sufficient for students to acquire certain concepts and skills such as, for example, estimation skills, but they should also get a feeling for situations and opportunities to use those skills; and, moreover, to become inclined to do so whenever appropriate. The acquisition of this disposition—especially the sensitivity and inclination aspects of it—requires extensive experience with the different categories of aptitude in a large variety of situations. As such, the disposition cannot be directly taught, but has to develop over a rather extensive period of time.

However, students' inclination and sensitivity to use available knowledge and skills can be blocked by emotional barriers. Taking this into account, our theoretical framework of a mathematical disposition can be complemented by Boekaerts' (1993) *affective learning process* model. According to this model, students confronted with a learning task develop either a learning intention or a coping intention, depending on how they perceive and experience the learning situation and the task demands. When positive expectations and feelings prevail, a learning intention develops; students are primarily oriented toward learning, and this leads to activity in the so-called mastery mode. In contrast, negative expectations and feelings generate a coping intention; students are not primarily concerned about learning, but about restoring their well-being, and this leads to coping activity. If students regain feelings of well-being, the appraisal of the situation can change and induce the development of a learning intention. Lehtinen, Vauras, Salonen, Olkinuora, and Kinnunen (1995) present a similar model, which contrasts three types of coping strategies—task oriented (which is in the mastery mode), ego defensive, and social dependence.

Linking the notion of a mathematical disposition to these models involves a move toward a more comprehensive and integrated conception of the cognitive and affective aspects of learning processes. Pursuing such an integration

represents a major challenge that faces present-day research on learning and instruction (see Shuell, 1992). Of course, our theoretical framework should be elaborated and validated in continued empirical work.

Designing powerful learning environments

Taking into account the preceding answer to the question of what should be learned to foster cognitive development, we are faced with the task of elaborating a set of inquiry-based principles for designing learning environments that are conducive to the acquisition of the intended mathematical disposition. In this respect, it is important to take into account the empirically supported characteristics of (effective) learning processes that have emerged from recent research on learning and instruction in general, and from the learning and teaching of mathematics in particular. After a brief overview of several such characteristics, three basic guidelines for the design of powerful learning environments are outlined. Finally, as an illustration, the main principles underlying Realistic Mathematics Education (RME), developed at the Freudenthal Institute in the Netherlands, are described.

Some major features of effective learning processes

Without trying to be exhaustive, a series of major characteristics of effective acquisition processes can be summarized in the following definition of learning: It is a constructive, cumulative, self-regulated, goal-oriented, situated, collaborative, and individually different process of knowledge and meaning building (see, e.g., Brown, Collins, & Duguid, 1989; Cobb, 1994; Shuell, 1992).

Learning is constructive (Cobb, 1994; De Corte, 1990; Glaser, 1991)

This overarching characteristic indicates that learners are not passive recipients of information, but that they construct their own knowledge and skill. Although there are differences along the continuum from radical to realistic constructivism (Cobb, 1994), the view certainly implies that acquiring new knowledge and skills is an active process, in the sense that it requires cognitive processing from the learner (Shuell, 1986). Referring to Salomon and Globerson (1987), one can say that effective learning is a mindful and effort-demanding activity.

Learning is cumulative (Dochy, 1992; Shuell, 1992; Vosniadou, 1992)

This refers to the crucial role of informal as well as formal prior knowledge for future learning. In fact, this feature is implied in the constructivist view of learning: It is on the basis of what they already know and can do that students actively process new information they encounter, and, as a consequence, derive new meanings and acquire new skills.

Learning is self-regulated (De Jong, 1992; Shuell, 1992; Simons, 1989; Vermunt, 1992)

This feature represents the metacognitive aspect of effective learning, especially the managing and monitoring activities of the student. According to Simons (1989), this involves "being able to prepare one's own learning, to take the necessary steps to learn, to regulate learning, to provide for one's own feedback and performance judgements and to keep oneself concentrated and motivated" (p. 16). The more learning becomes self-regulated, the more students can take control over their own learning; correlatively, they become less dependent on instructional support for performing this regulatory activity.

Learning is goal oriented (Bereiter & Scardamalia, 1989; Shuell, 1992)

Although learning also occurs incidently, effective and meaningful learning is facilitated by an explicit awareness of and orientation toward a goal. Taking into account its constructive and self-regulated nature, it is plausible to assume that learning is most productive when students determine and state their own goals. However, learning can also be successful when predefined objectives are put forward by a teacher, a textbook, a computer program, and so on, on the condition, however, that those goals are endorsed and adopted by the students themselves.

Learning is situated (Brown et al., 1989; Greeno, 1991a; Vygotsky, 1978)

In reaction to the view that knowledge acquisition is more or less a purely cognitive process that takes place inside the head and consists of the construction of mental representations, this characteristic stresses that learning essentially occurs in interaction with social and cultural context and artifacts, and especially through participation in cultural activities and practices.

Learning is cooperative (Brown et al., 1989; Vygotsky, 1978)

Because participation in social practices is an essential aspect of situated learning, it also implies the cooperative nature of productive learning. The view of learning as a social process is at present also central in the conception of most constructivists; it accounts for the fact that, notwithstanding the almost idiosyncratic processes of knowledge building, learners nevertheless acquire common concepts and skills. For example, Wood, Cobb, and Yackel (1991) considered social interaction essential for mathematics learning, with individual knowledge construction occurring throughout processes of interaction, negotiation, and cooperation. The impact of social interaction on knowledge acquisition and cognitive development is also supported by a substantial body of developmental research (see, e.g., Perret-Clermont & Schubauer-Leoni, 1989).

Learning is individually different (Ackerman, Sternberg, & Glaser, 1989; Entwistle, 1987; Marton, Dall'Alba, & Beaty, 1993; Snow & Swanson, 1992)

The outcomes and the processes of learning vary among students because of individual differences in a diversity of aptitudes that are relevant for learning, such as learning potential, prior knowledge, approaches to and conceptions of learning, interest, self-efficacy, self-worth, and so on. Individual differences in those aptitudes account for quantitative as well as qualitative variations in learning between students.

Design principles for powerful learning environments

In line with these characteristics of effective acquisition processes, and taking into account the idea of a mathematical disposition as the educational goal, the following principles can be put forward as guidelines for designing powerful learning environments (i.e., situations that can elicit in students the appropriate learning activities for achieving the intended outcomes; for a more detailed discussion, see De Corte, Greer, & Verschaffel, in press).

1. Learning environments should support the constructive, cumulative, goal-oriented acquisition processes in students. This also indicates that such environments must be designed to develop and enhance more active learning strategies in passive learners. In this respect, it is important to stress that conceiving learning as an active process does not, however, imply that students' construction of their knowledge and skills cannot be mediated through appropriate interventions and guidance by teachers, peers, and educational media such as modeling, coaching, and scaffolding (Collins, Brown, & Newman, 1989). In other words, a powerful learning environment is characterized by a good balance between discovery learning and personal exploration on the one hand and systematic instruction and guidance on the other.
2. Learning environments should foster students' self-regulation of their learning processes. This implies that external regulation of knowledge and skill acquisition in the form of systematic interventions should be gradually removed so that students become agents of their own learning (see Plowden, 1967). In other words, the balance between external and internal regulation will vary during students' learning history in the sense that progressively the share of self-regulation grows as explicit instructional support fades out.
3. Students' constructive learning activities should preferably be embedded in contexts that are rich in cultural resources, artifacts, and learning materials that offer ample opportunities for social interaction, and that are representative of the kind of tasks and problems to which the learners will

have to apply their knowledge and skills in the future. The acquisition of the disposition to develop good thinking and problem solving, especially the inclination and sensitivity aspects of this disposition, will require extensive experience and practice with the different categories of knowledge and skills in a large variety of situations.

4. Learning environments should allow for the flexible adaptation of the instructional support, especially the balance between self-discovery and direct instruction, or between self-regulation and external regulation, to take into account the individual differences among learners in cognitive aptitudes as well as in affective and motivational characteristics. In addition, the important impact of motivational characteristics on learning activities and outcomes point to the necessity of alternating instructional interventions with emotional support, depending on whether the individual student is in the learning or in the coping mode (Boekaerts, 1993).
5. Because domain-specific knowledge, heuristic methods, and metacognitive knowledge and strategies play a complementary role in competent learning, thinking, and problem solving, learning environments should create possibilities to acquire general learning and thinking skills embedded in the different subject-matter domains.

There is no doubt that these guiding principles need to be validated thoroughly in future intervention studies. Nevertheless, a number of success stories that embody those principles to some degree have already reported initial supporting empirical evidence. Examples of such success stories are Lampert's (1986) collaborative teaching of multiplication, anchored instruction designed by the Cognition and Technology Group at Vanderbilt (1993), Schoenfeld's (1985) heuristic teaching of mathematical problem solving, Cobb's second-grade mathematics project (Cobb et al., 1988), and RME (Streefland, 1991b; Treffers, 1987). These examples represent a rather radical departure from traditional, weak classroom environments; it is based on the view that mathematics learning is a highly individual activity, consisting mainly in absorbing and memorizing a fixed body of decontextualized and fragmented knowledge and procedural skills transmitted by the teacher. Taking into account the European flavor of this issue, only RME is discussed here.

Realistic Mathematics Education (RME)

RME clearly embodies a number of the features of effective acquisition processes and guiding principles for powerful learning environments described earlier. This is remarkable for two reasons. First, RME originates primarily from a mathematics education perspective and not directly from a psychological approach to mathematics learning and teaching. Second, and more importantly, RME was founded in the early 1970s in reaction to the then dominating mechanistic approach to mathematics instruction in the Netherlands, and

thus emerged many years before most of the research on learning and instruction reviewed earlier.

In contrast to the still-prevailing view in educational practice of mathematics as a universal, formal system of concepts and rules that has to be transmitted as precisely as possible from one generation to the next, RME conceives mathematics as a human activity focused on problem solving and construction of meaning. Therefore, learning mathematics essentially consists of doing mathematics, or mathematizing. This view is based on Freudenthal's (1983) so-called didactical phenomenology. He argued that reality serves not only as a domain of application of knowledge, but in the first place as a source that enables the learners to constitute so-called mental objects (i.e., intuitive notions that precede concept attainment). This also implies that the learning environment has to be adaptive to the learners to facilitate the intended process of reinvention of mathematics knowledge.

Starting from this fundamental conception of doing mathematics, the design of realistic learning environments is guided by the following five interrelated principles:

1. The major role of context problems, serving as a source for the construction of mathematical concepts, but also as a field of their application;
2. The extensive use of models as tools or scaffolds to facilitate progression toward higher levels of abstraction;
3. The important contribution of children's own constructions and productions as a starting point for reflection;
4. The importance of interaction and cooperation for learning; and
5. The intertwining of learning strands.

In the remainder of this section, each of these principles is briefly discussed (for further details, see Treffers, 1987, 1991; Treffers & Goffree, 1985).

Context problems

The first principle underlying RME is that learners do not absorb concepts and procedures passively, but that they actively construct their mathematical knowledge and skills starting from the exploration of socalled context problems, using their own informal knowledge and working methods. *Context problems* are mathematical problems that are presented within a broader framework of real-life situations with which children are familiar, or through motivating stories derived from the world of fantasy. They can be presented in a variety of formats, such as a word problem, a game, a drawing, a newspaper clipping, a graph, or a combination of these kinds of information. Such meaningful context problems offer a concrete orientation for the acquisition of a new concept or skill, and they allow students to invoke and use their prior knowledge.

For example, it has been shown that third graders can discover progressively a procedure for long division that comes close to the standard division algorithm, starting from exploring context problems like "The PTA meeting at our school will be attended by 81 parents. Six parents can be seated at one table. How many tables will be needed?" and "How many pots of coffee will have to be made for the parents? One pot serves seven cups of coffee, and each parent will be offered one cup."

A variety of solution methods were initially proposed in a class of 17 third graders, ranging from very simple (e.g., repetitive addition) to more sophisticated ones (e.g., using 10×6 as a starting point). After comparing and discussing the distinct strategies in the class most children rather quickly switched to the more efficient "ten times" method for solving the problem concerning the number of pots of coffee. This happened spontaneously, in the sense that the teacher had not given any hint or suggestion to do so.

Through progressive schematization the class invented the following long-division scheme:

7/81	
70	10 pots
11	
7	1 pot
4	
4	(1 pot)
0	12 pots of coffee

Besides their important role of serving as a source for meaningful learning of new concepts and procedures, context problems are also used as a domain of application of the acquired mathematical knowledge and skills.

Models as scaffolds to facilitate abstraction

Acquiring mathematical concepts and skills is a long-term process involving a progression toward increasing levels of abstraction. For instance, the aforementioned progressive schematization in learning long division requires an increase in the level of abstraction in the procedures that students use. RME employs a variety of mathematical tools and models to scaffold the transition from the concrete, intuitive level to the abstract, formal level of mathematics. Manipulatives, visual and situation models, diagrams, schemes, and symbols can fulfill this bridging function; specific examples are the empty or unstructured number line, the abacus, the arrow diagram, and the rectangle model. It is important to note that in RME the notion of *level of abstraction* refers to the degree of closeness to context problems. The low level remains close to the context problem and allows children to use informal knowledge and strategies; at higher levels, children work within the formal system of mathematics, requiring the application of abstract and formalized procedures.

Encouraging students' own productions and reflection

The role of children's free constructions and productions for mathematics learning was implied earlier, namely that students actively construct their knowledge and skills starting from the exploration of context problems. In addition, this guideline stresses the importance of free productions as a starting point for reflection. Indeed, by creating opportunities in the classroom for considering and discussing children's productions, the teacher induces reflection, which is a major vehicle in promoting the attainment of higher levels of abstraction during mathematics learning. For instance, it was through reflection on students' own informal methods for solving the PTA meeting problem that the class invented a scheme for long division that went in the direction of the standard division algorithm.

A second example of this principle relates to the following task (Streefland, 1988, p. 8):

> Invent stories that involve dividing 6,394 by 12, such that the result is, respectively:
>
> 532
> 533
> 532 remainder 10
> 532.84 remainder 4
> 532.833333
> about 530.

Reflection on students' productions in response to this task can promote their understanding of the operation of division and make them aware that the meaning of the remainder of a division can vary, depending on the context or situation of the problem (see also Gravemeijer, Van den Heuvel, & Streefland, 1990).

Interactive and cooperative learning

It is obvious from the preceding discussion that, in RME, learning mathematics is not considered a purely solitary enterprise, but rather as an activity that takes place in and is facilitated by a social context. Social interaction and cooperation are considered crucial because of the importance in learning and doing mathematics of exchanging and negotiating ideas, comparing solution methods, and discussing arguments. Of special significance in this respect is that interaction and cooperation mobilize reflection. Consequently, in RME whole-class instruction and individual work are combined with cooperative learning in small groups and classroom discussion. To guarantee the quality of this learning through interaction, the role of the teacher is essential: for instance, eliciting explicit description and justification of

children's own solutions, stimulating comparison of and reflection on different approaches and strategies, and encouraging the search for more efficient solution methods.

Intertwining of learning strands

When I discussed the role of domain-specific knowledge in thinking and problem solving (Acquiring a Mathematical Disposition section), I stressed the importance of building up a well-organized, coherent, and flexibly accessible knowledge base in which subject-matter elements such as concepts and rules are closely interconnected. This is exactly what this principle implies. It derives from the phenomenological basis of RME; indeed, the real phenomena that underlie the mathematical concepts, rules, and structures in the different learning strands are interrelated in manifold ways, and constitute an organized and meaningful whole. For instance, instruction should explicitly link division to the other basic operations, mental arithmetic to written computation, proportions to fractions, and measurement to geometry.

The preceding discussion obviously shows that RME fits in well with the research findings and ideas presented earlier, namely the acquisition of a mathematical disposition as the overarching objective, the constructive nature of learning, taking children's informal knowledge and strategies as the starting point for engaging them in mathematical activity, embedding learning in realistic contexts, and the importance of interaction and cooperation for effective learning. But is there empirical evidence to support the educational benefits and value of RME? And what about its implementation in classroom practice?

Well-designed experiments and evaluation research relating to RME are rather scarce. In one study, instruction of long division according to the RME approach in an experimental class resulted in better performance reached in about half the time as compared to a control group with traditional teaching (Treffers, 1987). Streefland (1991a) also obtained promising results in support of the RME approach to teaching fractions. In addition, there are quite a number of anecdotal studies that report qualitative data showing positive results (see, e.g., Streefland, 1991b; Van den Brink, 1991).

A major achievement of the Freudenthal Institute is certainly the production of a plan for a national curriculum for mathematics education in The Netherlands (Treffers, De Moor, & Feijs, 1989) that embodies the RME approach. This document is a counterpart of the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) in the United States. In addition, the Freudenthal Institute has elaborated lesson plans for certain aspects of the curriculum, such as written multiplication and division, fractions, and so on. Furthermore, RME has substantially influenced Dutch textbooks for mathematics education in the primary schools (see De Corte, Greer, & Verschaffel, in press). But a recent study by

Gravemeijer et al. (1993) has confirmed the well-known phenomenon that the availability of RME-based textbooks does not guarantee that teachers using those handbooks will implement the approach appropriately. Therefore, substantial efforts have been undertaken in recent years to introduce the RME approach more systematically in preservice as well as in-service teacher training.

Remaining research issues

Although it has proved possible to identify a series of principles from research on learning and instruction from which to design powerful learning environments, and although RME and several related projects (mentioned earlier) exemplify and to some extent validate those principles, there nevertheless remain major issues for continued research.

First, the principles for the design of powerful learning environments outlined in this article need further elaboration and more thorough validation in future intervention studies. This constitutes a challenging joint task for scholars in the domains of mathematics education and the psychology of mathematics learning and instruction, in cooperation with interested expert practitioners. But there is also a strong need for thorough theory-oriented research that aims at a better understanding and fine-grained analysis of the constructive learning processes that the new kind of learning environments elicit in children; of the precise nature of the knowledge, skills, attitudes and beliefs that they acquire; and of the critical dimensions that can account for the power and efficacy of these environments.

Second, for this kind of research, it is also necessary to develop a methodology for the construction of new forms of assessment that can tap the relevant aspects of students' learning activities and outcomes, and that are sensitive to instructional components of learning environments. A promising approach in pursuing these research objectives seems to be the application of so-called design experiments (Brown, 1992; Collins, 1992) in which investigators, in close cooperation with practitioners, elaborate and evaluate innovative teaching and learning environments and, at the same time, use those environments as a "workbench" for carrying out theory-oriented research.

Third, a problem of utmost importance, emerging clearly from the RME project, relates to the appropriate implementation of the new approaches to mathematics learning and teaching. Here one is again confronted with the Achilles' heel of educational innovation and improvement: preservice and in-service teacher training. One should realize from the outset that solving this problem is difficult and time consuming, because it is not just a matter of acquiring a new set of teaching techniques and skills, but instead requires fundamental changes in people's conceptions and beliefs about (mathematics) learning and teaching.

References

Ackerman, P. L., Sternberg, R. J., & Glaser, R. (Eds.). (1989). *Learning and individual differences: Advances in theory and research*. New York: Freeman.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361–392). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Boekaerts, M. (1993). Being concerned with well-being and with learning. *Educational Psychologist*, *28*, 149–167.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in evaluating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*, 141–178.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In P. H. Mussen, J. H. Flavell, & E. M. Markman (Eds.), *Child psychology: Vol. 3. Cognitive development* (pp. 77–166). New York: Wiley.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Development Psychology*, *27*, 703–722.

Chi, M. T., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cobb, P. (1994). Constructivism and learning. In T. Husen & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed., pp. 1040–1052). Oxford, England: Pergamon.

Cobb, P., Wood, T., Yackel, E., McNeal, G., Preston, M., & Wheatley, G. (1988). *The Purdue problem-centered mathematics curriculum: Revised*. West Lafayette, IN: Purdue University, School Mathematics and Science Center.

Cognition and Technology Group at Vanderbilt. (1993). Anchored instruction and situated cognition revisited. *Educational Technology*, *33*(3), 52–70.

Coleman, J. S., Campbell, E. Q., Hobson, C., McPartland, J., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (NATO ASI Series F: Computer and Systems Sciences, Vol. 96, pp. 15–22). Berlin: Springer-Verlag.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

De Corte, E. (1990). Acquiring and teaching cognitive skills: A state-of-the-art of theory and research. In P. J. D. Drenth, J. A. Sergeant, & R. J. Takens (Eds.), *European perspectives in psychology* (Vol. 1, pp. 237–263). London: Wiley.

De Corte, E., Greer, B., & Verschaffel, L. (in press). Mathematics learning and teaching. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology*. New York: Macmillan.

De Corte, E., & Verschaffel, L. (1987). Children's problem solving skills and processes with respect to elementary arithmetic word problems. In E. De Corte,

H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction. European research in an international context* (Vol. 1, pp. 297–308). Oxford: Pergamon Press.

De Corte, E., Verschaffel, L., & Van Coillie, V. (1988). Influence of number size, problem structure, and response mode on children's solutions of multiplication problems. *Journal of Mathematical Behavior*, *7*, 197–216.

De Jong, F. P. C. M. (1992). *Zelfstandig leren. Regulatie van het leeproces en leren reguleren: Een procesbenadering* [Independent learning. Regulation of the learning process and learning to regulate: A process approach]. Tilburg, The Netherlands: Katholieke Universiteit Brabant.

Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning*. Utrecht, The Netherlands: Lemma.

Dweck, C. S., & Elliott, E. S. (1983). Achievement motivation. In P. H. Mussen (Ed.), *Handbook of child psychology* (Vol. 4, pp. 643–692). New York: Wiley.

Entwistle, N. J. (1987). *Understanding classroom learning*. London: Hodder & Stoughton.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, Holland: Reidel.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan.

Glaser, R. (1991). The maturing of the relationship between the science of learning and cognition and educational practice. *Learning and Instruction*, *1*, 129–144.

Gravemeijer, K., Van de Heuvel, M., & Streefland, L. (1990). *Context, free productions, tests and geometry in realistic mathematics education*. Utrecht, The Netherlands: University of Utrecht, Research Group for Mathematical Education and Educational Computer Centre.

Gravemeijer, K., Van den Heuvel-Panhuizen, M., Van Donselaar, G., Reusink, N., Streefland, L., Vermeulen, W., Te Woerd, E., & Van der Ploeg, D. (1993). *Methoden in het reken/wiskunde onderwijs, een rijke contextvoor vergelijkend onderzoek* [Teaching methods in mathematics education, a rich context for comparative research]. Utrecht, The Netherlands: DB-beta, Centrum voor beta-Didactiek.

Greeno, J. G. (1991a). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education*, *22*, 170–218.

Greeno, J. G. (1991b). A view of mathematical problem solving in school. In M. U. Smith (Ed.), *Toward a unified theory of problem solving: Views from the content domains* (pp. 69–98). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Greer, B. (1992). Multiplication and division as models of situations. In D. A. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 276–295). New York: Macmillan.

Helmke, A. (1992, July). *The development of children's attitudes towards learning in the elementary school: A longitudinal study*. Paper presented at the 25th International Congress of Psychology, Brussels, Belgium.

Husen, T., & Tuijnman, A. (1991). The contribution of formal schooling to the increase in intellectual capital. *Educational Researcher*, *20*, 17–25.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Mitchelson, S. (1972). *Inequality: A reassessment of the effects of family and schooling in America*. New York: Basic Books.

Lampert, M. (1986). Knowing, doing, and teaching multiplication. *Cognition and Instruction*, *3*, 305–342.

Lehtinen, E., Vauras, M., Salonen, P., Olkinuora, E., & Kinnunen, R. (1995/this issue). Long-term development of learning activity: Motivational, cognitive, and social interaction. *Educational Psychologist*, *30*, 21–35.

Marton, F., Dall'Alba, G., & Beaty, E. (1993). Conceptions of learning. *International Journal of Educational Research*, *19*, 277–300.

McLeod, D. B. (1990). Information-processing theories and mathematics learning: The role of affect. *International Journal of Educational Research*, *14*, 13–29.

McLeod, D. B., & Adams, V. M. (1989). *Affect and mathematical problem solving. A new perspective*. New York: Springer-Verlag.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematic*s. Reston, VA: National Council of Teachers of Mathematics.

Nelissen, J. M. C. (1987). *Kinderen leren wiskunde. Een studie over constructie en reflectie in het basisonderwijs* [Children learning mathematics. A study on construction and reflection in elementary school children]. Gorinchem, The Netherlands: Uitgeverij De Ruiter.

Overtoom, R. (1991). *Informatieverwerking door hoogbegaafde leerlingen bij het oplossen van wiskundeproblemen* [Information processing by gifted students in solving mathematical problems]. De Lier, The Netherlands: Academisch Boeken Centrum.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill–Palmer Quarterly*, *39*, 1–21.

Perret-Clermont, A., & Schubauer-Leoni, M. (Eds.). (1989). Social factors in learning and teaching. *International Journal of Educational Research*, *13*, 573–684.

Plowden, B. H. (1967). *Children and their primary schools: A report of the Central Advisory Council for Education*. London: Her Majesty's Stationery Office.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. *International Journal of Educational Research*, *11*, 326–637.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189–215). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist*, *23*, 145–166.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics learning and teaching* (pp. 334–370). New York: Macmillan.

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research*, *56*, 411–436.

Shuell, T. J. (1992). Designing instructional computing systems for meaningful learning. In M. Jones & P. H. Winne (Eds.), *Adaptive learning environments: Foundations and frontiers* (NATO ASI Series F: Computer and Systems Sciences, Vol. 85, pp. 19–54). Berlin: Springer-Verlag.

Simons, P. R. J. (1989). Learning to learn. In P. Span, E. De Corte, & B. van Hout-Wolters (Eds.), *Onderwijsleerprocessen: Strategieën voor de verwerking van informatie* (pp. 15–25). Amsterdam: Swets & Zeitlinger.

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, *27*, 5–32.

Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology*, *43*, 583–626.

Streefland, L. (1988). Reconstructive learning. In A. Borbas (Ed.), *Proceedings of the Twelfth International Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 75–91). Veszprem, Hungary: OOK Printing House.

Streefland, L. (1991a). *Fractions in Realistic Mathematics Education: A paradigm of developmental research*. Dordrecht, The Netherlands: Kluwer.

Streefland, L. (Ed.). (1991b). *Realistic Mathematics Education in primary school. On the occasion of the opening of the Freudenthal Institute*. Utrecht, The Netherlands: Utrecht University, Freudenthal Institute.

Treffers, A. (1987). *Three dimensions. A model of goal and theory description in mathematics instruction. The Wiskobas Project*. Dordrecht, The Netherlands: D. Reidel.

Treffers, A. (1991). Didactical background of a mathematics program for primary education. In L. Streefland (Ed.), *Realistic Mathematics Education in primary school. On the occasion of the opening of the Freudenthal Institute* (pp. 21–56). Utrecht, The Netherlands: Utrecht University, Freudenthal Institute.

Treffers, A., De Moor, E., & Feijs, E. (1989). *Proeve van een nationaal programma voor het rekenwiskundeonderwijs op de basisschool. Deel 1: Overzicht leerdoelen* [Plan for a national curriculum for mathematics education at the primary school. Part 1: Overview of objectives]. Tilburg, The Netherlands: Zwijsen.

Treffers, A., & Goffree, F. (1985). Rational analysis of Realistic Mathematics Education: The Wiskobas program. In L. Streefland (Ed.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education: Vol. 2. Plenary addresses and invited papers* (pp. 97–121). Utrecht, The Netherlands: University of Utrecht, Research Group on Mathematics Education and Educational Computer Centre.

Van den Brink, J. (1991). Didactic constructivism. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 195–227). Dordrecht, The Netherlands: Kluwer.

Vermunt, J. D. H. M. (1992). *Leerstijlen en sturen van leerprocessen in het hoger onderwijs: Naar procesgerichte instructie in zelfstandig denken* [Learning styles and regulation of learning processes in higher education: Toward process-oriented instruction in independent thinking]. Amsterdam, The Netherlands: Swets & Zeitlinger.

Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology: An International Journal*, *41*, 347–357.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wood, T., Cobb, P., & Yackel, E. (1991). Change in teaching mathematics. *American Educational Research Journal*, *28*, 587–616.

72

SOCIOMATHEMATICAL NORMS, ARGUMENTATION, AND AUTONOMY IN MATHEMATICS

E. Yackel and P. Cobb

This paper sets forth a way of interpreting mathematics classrooms that aims to account for how students develop mathematical beliefs and values and, consequently, how they become intellectually autonomous in mathematics. To do so, we advance the notion of *sociomathematical* norms, that is, normative aspects of mathematical discussions that are specific to students' mathematical activity. The explication of sociomathematical norms extends our previous work on general classroom social norms that sustain inquiry-based discussion and argumentation. Episodes from a second-grade classroom where mathematics instruction generally followed an inquiry tradition are used to clarify the processes by which sociomathematical norms are interactively constituted and to illustrate how these norms regulate mathematical argumentation and influence learning opportunities for both the students and the teacher. In doing so, we both clarify how students develop a mathematical disposition and account for students' development of increasing intellectual autonomy in mathematics. In the process, the teacher's role as a representative of the mathematical community is elaborated.

For the past several years, we have been engaged in a research and development project at the elementary school level that has both pragmatic and theoretical goals. On one hand, we wish to support teachers as they establish classroom environments that facilitate students' mathematical conceptual development. On the other hand, we wish to investigate children's mathematical learning in the classroom. The latter involves developing perspectives that are useful for interpreting and attempting to make sense of the complexity of classroom life. The purpose of this paper is to set forth a way of interpreting classroom life that aims to account for how students develop

Source: *Journal for Research in Mathematics Education*, 1996, 27(4), 458–477.

specific mathematical beliefs and values and, consequently, how they become intellectually autonomous in mathematics, that is, how they come to develop a mathematical disposition (National Council of Teachers of Mathematics, 1991). To that end, we focus on classroom norms that we call sociomathematical norms. These norms are distinct from general classroom social norms in that they are specific to the mathematical aspects of students' activity. As a means of introducing and elaborating the theoretical discussion in this paper, we present episodes from a classroom that we have studied extensively. The episodes have been selected for their clarifying and explanatory power and are not meant to be exemplary or reflect ideal classroom practice.

There is a reflexive relationship between developing theoretical perspectives and making sense of particular events and situations. The analysis of the particular constitutes occasions to reconsider what needs to be explained and to revise explanatory constructs. Conversely, the selection of particulars to consider reflects one's theoretical orientation. Thus, particular events empirically ground theoretical constructs, and theoretical constructs influence the interpretation of particular events (Erickson, 1986). This interdependence between theory and practice is reflected throughout this paper.

Theoretical perspective

Our theoretical perspective is derived from constructivism (von Glasersfeld, 1984), symbolic interactionism (Blumer, 1969), and ethnomethodology (Leiter, 1980; Mehan & Wood, 1975). We began the project intending to focus on learning primarily from a cognitive perspective, with constructivism as a guiding framework. However, as we attempted to make sense of our experiences in the classroom, it was apparent that we needed to broaden our interpretative stance by developing a sociological perspective on mathematical activity. For this purpose, we drew on constructs derived from symbolic interactionism (Bauersfeld, Krummheuer, & Voigt, 1988; Blumer, 1969; Voigt, 1985, 1989) and ethnomethodology (Krummheuer, 1983; Mehan & Wood, 1975). We were then able to account for and explicate the development of general classroom social norms. These same constructs proved critical to our development of the notion of sociomathematical norms. As will be seen throughout, constructs that proved particularly relevant are the interactive constitution of meaning, from symbolic interactionism, and reflexivity, from ethnomethodology. A detailed discussion of the coordination of psychological and sociological perspectives is beyond the scope of this paper and can be found in Cobb and Bauersfeld (1995).

Bauersfeld (1988) and Voigt (1992) have elaborated the relevance of interactionist perspectives for mathematics education research. A basic assumption of interactionism is that cultural and social processes are integral to mathematical activity (Voigt, 1995). This view, which is increasingly

accepted by the mathematics education community (Cobb, 1990; Eisenhart, 1988; Greeno, 1991; Resnick, 1989; Richards, 1991), is stated succinctly by Bauersfeld (1993).

> [T]he understanding of learning and teaching mathematics . . . support[s] a model of participating in a culture rather than a model of transmitting knowledge. Participating in the processes of a mathematics classroom is participating in a culture of using mathematics, or better: a culture of mathematizing as a practice. The many skills, which an observer can identify and will take as the main performance of the culture, form the procedural surface only. These are the bricks for the building, but the design for the house of mathematizing is processed on another level. As it is with cultures, the core of what is learned through participation is *when* to do what and *how* to do it. Knowledge (in a narrow sense) will be for nothing once the user cannot identify the adequateness of a situation for use. Knowledge, also, will not be of much help, if the learner is unable to flexibly relate and transform the necessary elements of knowing into his/her actual situation. This is to say, the core effects as emerging from the participation in the culture of a mathematics classroom will appear on the metalevel mainly and are "learned" indirectly. (p. 4)

In this view, the development of individuals' reasoning and sense-making processes cannot be separated from their participation in the interactive constitution of taken-as-shared mathematical meanings.

Voigt (1992) argues that, of the various theoretical approaches to social interaction, the symbolic interactionist approach is particularly useful when studying children's learning in inquiry mathematics classrooms because it emphasizes the individual's sense-making processes as well as the social processes. Thus, rather than attempting to deduce an individual's learning from social and cultural processes or vice versa, it treats "subjective ideas as becoming compatible with culture and with intersubjective knowledge like mathematics" (Voigt, 1992, p. 11). Individuals are therefore seen to develop their personal understandings as they participate in negotiating classroom norms, including those that are specific to mathematics.

As we will demonstrate, the construct of reflexivity from ethnomethodology (Leiter, 1980; Mehan & Wood, 1975) is especially useful for clarifying how sociomathematical norms and goals and beliefs about mathematical activity and learning evolve together as a dynamic system. Methodologically, both general social norms and sociomathematical norms are inferred by identifying regularities in patterns of social interaction. With regard to sociomathematical norms, what becomes mathematically normative in a classroom is constrained by the current goals, beliefs, suppositions, and assumptions of the classroom

participants. At the same time these goals and largely implicit understandings are themselves influenced by what is legitimized as acceptable mathematical activity. It is in this sense that we say sociomathematical norms and goals and beliefs about mathematical activity and learning are reflexively related.

Social and sociomathematical norms

In the course of our work, we have collaborated with a group of second- and third-grade teachers to help them radically revise the way they teach mathematics. Instruction in project classrooms typically consists of teacher-led discussions of problems posed in a whole-class setting, collaborative small-group work between pairs of children, and follow-up whole-class discussions where children explain and justify the interpretations and solutions they develop during small-group work. The instructional tasks and the instructional strategies used in project classrooms have been developed during several yearlong classroom teaching experiments. In general, the approach we have taken reflects the view that mathematical learning is both a process of active individual construction (von Glasersfeld, 1984) and a process of acculturation into the mathematical practices of wider society (Bauersfeld, 1993).

Our prior research has included analyzing the process by which teachers initiate and guide the development of social norms that sustain classroom microcultures characterized by explanation, justification, and argumentation (Cobb, Yackel, & Wood, 1989; Yackel, Cobb, & Wood, 1991). Norms of this type are, however, general classroom social norms that apply to any subject matter area and are not unique to mathematics. For example, ideally students should challenge others' thinking and justify their own interpretations in science or literature classes as well as in mathematics. In this paper we extend our previous work on general classroom norms by focusing on normative aspects of mathematics discussions specific to students' mathematical activity. To clarify this distinction, we will speak of *sociomathematical* norms rather than social norms. For example, normative understandings of what counts as mathematically different, mathematically sophisticated, mathematically efficient, and mathematically elegant in a classroom are sociomathematical norms. Similarly, what counts as an acceptable mathematical explanation and justification is a sociomathematical norm. To further clarify the subtle distinction between social norms and sociomathematical norms we offer the following examples. The understanding that students are expected to explain their solutions and their ways of thinking is a social norm, whereas the understanding of what counts as an acceptable mathematical explanation is a sociomathematical norm. Likewise, the understanding that when discussing a problem students should offer solutions different from those already contributed is a social norm, whereas the understanding of what constitutes mathematical difference is a sociomathematical norm.

In this paper we first document the processes by which the sociomathematical norms of *mathematical difference* and *mathematical sophistication* are established. Next, we illustrate how these sociomathematical norms regulate mathematical argumentation and influence the learning opportunities for both the students and the teacher. We then consider how the teacher and students interactively constitute what counts as an *acceptable mathematical explanation and justification*. In the process, we clarify how the teacher can serve as a representative of the mathematical community in classrooms where students develop their own personally meaningful ways of knowing.

Issues concerning what counts as different, sophisticated, efficient, and elegant solutions involve a taken-as-shared sense of *when* it is appropriate to contribute to a discussion. In contrast, the sociomathematical norm of what counts as an acceptable explanation and justification deals with the actual *process* by which students contribute. Because teachers with whom we collaborated were attempting to establish inquiry mathematics traditions in their classrooms, acceptable explanations and justifications had to involve described actions on mathematical objects rather than procedural instructions (Cobb, Wood, Yackel, & McNeal, 1992). For example, describing manipulation of numerals per se would not be acceptable. On the other hand, it was not sufficient for a student to merely describe personally real mathematical actions. Crucially, to be acceptable, other students had to be able to interpret the explanation in terms of actions on mathematical objects that were experientially real to them. Thus, the currently taken-as-shared basis for mathematical communication served as the backdrop against which students explained and justified their thinking. Conversely, it was by means of mathematical argumentation that this constraining background reality itself evolved. We will therefore argue that the process of argumentation and the taken-as-shared basis for communication were reflexively related.

Further, we will argue that the construct of sociomathematical norms is pragmatically significant, in that it clarifies how students in classrooms that follow an inquiry tradition develop mathematical beliefs and values that are consistent with the current reform movement and how they become intellectually autonomous in mathematics. Therefore, in keeping with the purpose of this paper, we limit our discussion to classrooms that follow an inquiry tradition. Nevertheless, sociomathematical norms, such as what counts as an acceptable mathematical explanation and justification, are established in all classrooms regardless of instructional tradition.

To clarify the theoretical constructs developed in this paper, we have selected examples from a second-grade classroom in which we conducted a yearlong teaching experiment. Data from the teaching experiment include video recordings for all mathematics lessons for the entire school year and of individual interviews conducted with each student in the class at the beginning, middle, and end of the school year. Field notes and copies of students' written work are additional data sources.

The process of developing sociomathematical norms

As part of the process of guiding the development of a classroom atmosphere in which children are obliged to try to develop personally meaningful solutions that they can explain and justify, the teachers with whom we have worked regularly asked if anyone had solved a problem in a *different way*. It was while we were analyzing teachers' and students' interactions in these situations that the importance of sociomathematical norms, as opposed to general social norms, first became apparent. We will use the notion of *mathematical difference* to clarify and illustrate how sociomathematical norms are interactively constituted in the classroom.

In project classrooms, as in most mathematics classrooms, there were no pregiven criteria for what counted as a *different* solution. Instead, the meaning of what constituted *mathematical difference* was negotiated by each teacher and his or her students through their interaction. For their part, the teachers were themselves attempting to develop an inquiry form of practice. They did not have prior experience asking children to generate their own solution methods or explain their own thinking and, therefore, had little basis for anticipating methods the children would suggest. In the absence of predetermined criteria, the children had to offer solution methods without knowing in advance how they would be viewed by the teacher. Consequently, in responding to the teacher's requests for different solutions, the students were simultaneously learning what counts as mathematically different and helping to constitute what counts as mathematically different in their classroom. It is in this sense that we say the meaning of *mathematical difference* was interactively constituted by the teacher and the children. The teacher's responses and actions constrained the students' developing understanding of mathematical difference and the students' responses contributed to the teacher's developing understanding.

The following episode clarifies and illustrates how the teacher initiates the interactive constitution of mathematical difference.

Example 1: The number sentence 16 + 14 + 8 =____ has been posed as a mental computation activity.

Lemont: I added the two 1s out of the 16 and [the 14] . . . would be 20 . . . plus 6 plus 4 would equal another 10, and that was 30 plus 8 left would be 38.

Teacher: All right. Did anyone add a little *different*? Yes?

Ella: I said 16 plus 14 would be 30 . . . and add 8 more would be 38.

Teacher: Okay! Jose? *Different*?

Jose: I took two tens from the 14 and the 16 and that would be 20 . . . and then I added the 6 and the 4 that would be 30 . . . then I added the 8, that would be 38.

Teacher: Okay! It's almost similar to—(Addressing another student) Yes? *Different*? All right.

Here, the teacher's response to Jose suggests that he is working out for himself the meaning of *different*. However, because he does not elaborate for the students how Jose's solution is similar to those already given, the students are left to develop their own interpretations. The next two solutions offered by students are more inventive and are not questioned by the teacher.

Rodney: I took one off the 6 and put on the 14 and I had . . . 15 [and] 15 [would be] 30, and I had 8 would be 38.
Teacher: Yeah! Thirty-eight. Yes. *Different*?
Tonya: I added the 8 and the 4, that was 12. . . . So I said 12 plus 10, that would equal 22 . . . plus the other 10, that would be 30—and then I had 38.
Teacher: Okay! Dennis—*different*, Dennis?

By participating in exchanges such as this, the children learned that the teacher legitimized solutions that involved decomposing and recomposing numbers in differing ways but not those that were little more than restatements of previously given solutions. At the same time, the teacher furthered his pedagogical agenda by guiding the development of a taken-as-shared understanding of what was mathematically significant in such situations.

The next example further highlights the subtle and often implicit negotiation of the meaning. In this case, we see a student taking the initiative as he protests that a solution should not have been offered because, in his view, it was not different from one already given.

Example 2: The problem 78 – 53 =____ was written on the chalkboard and posed as a mental computation activity.

Dennis: I said, um, 7 and take away 50, that equals 20.
Teacher: All right.
Dennis: And then, then I took, I took 3 from that 8 and then that left 5.
Teacher: Okay. And how much did you get?
Dennis: 25. . . .
. . .
Teacher: Ella?
Ella: I said the 7, the 70, I said the 70 minus the 50 . . . I said the 20 and 8 plus 3, . . . Oh, I added, I said 8 minus the 3, that'd be 5.
Teacher: All right. It'd be what?
Ella: And that's 75 . . . I mean 25.
Dennis: (Protesting) Mr. K., that's the same thing I said.

Dennis's final comment serves two functions. With regard to the class discussion, it contributes to the negotiation of the meaning of *mathematical difference*. For the observer, it shows he understands that in this class it is not appropriate to offer an explanation that repeats a previously described decomposition and recombination of numbers. The notion of when it is appropriate to contribute to the discussion was taken as shared by at least some members of the class.

The preceding example clarifies that, in addition to regulating their participation in discussion, the sociomathematical norm of what constitutes mathematical difference supports higher-level cognitive activity. To respond as he did, Dennis had to compare his and Ella's solutions and judge the similarities and differences. In doing so, his solution became an object of his own reflection. In general, the teacher's requests for different solutions initiate a change in the setting from solving the problem to comparing solutions. In the latter setting the children's activity extends beyond listening to, and trying to make sense of, the explanations of others to attempting to identify similarities and differences among various solutions. Such reflective activity has the potential to contribute significantly to children's mathematical learning.

In the classroom studied, developing a taken-as-shared understanding of what counts as a *sophisticated solution* or an *efficient solution* was less explicit than an understanding of what counts as a *different solution*. For example, in this classroom the teacher rarely asked if anyone had a *more sophisticated* way or a *more efficient* way to solve a problem and never explicitly referred to one solution as better than another. Nevertheless, in any classroom, children are well aware of the asymmetry between the teacher's role and their role. The teacher necessarily represents the discipline of mathematics in the classroom (Voigt, 1995). Consequently, the teacher's reactions to a child's solution can be interpreted as an implicit indicator of how it is valued mathematically. For instance, in Example 1, many children may have interpreted the teacher's enthusiastic response ("Yeah!") following Rodney's solution as an indication that this solution was favored. However, because the issue did not become an explicit topic of conversation, the children were left to decide in what sense the solution was special. Events of this type are occasions for the children to infer what aspects of their mathematical activity the teacher values. In the process, the teacher both elaborates his own interpretative stance toward mathematics and inducts students into that stance.

The following episode, which occurred within the first few weeks of the school year, clarifies how mathematical discourse can advance as the teacher and students interactively constitute a taken-as-shared understanding of what is valued mathematically.

Example 3: The task is to figure out how many chips there are in a double-tens frame that has four red chips on the left frame and six green chips on

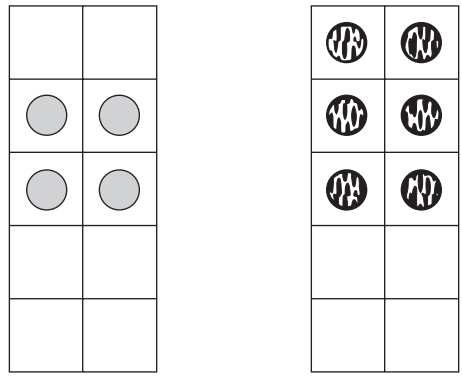


Figure 1 Double tens-frame task

the right frame (see Figure 1). The image was flashed on the overhead screen several times and then left off while the children figured out their solutions. The episode begins after several children have already given solutions that involve counting by ones.

Travonda: You could say, um, um, it's 6 on this side (pointing to the right frame) and take one from that side (pointing to the right frame) [and] put it on the red side and . . .
Teacher: *Listen to her!*
Travonda: And [you] would have 5 plus 5.
Teacher: *All right! Do you understand what she* [said]. *I like that!* She said (pointing to the screen) if we were to take one of these green and put it over here with, with the four [red chips] we'd have what?
Class: Five.
Teacher: Five. And this would leave five here (pointing to the right tens frame) and you could say 5 plus 5. *That's good.*

Even though the teacher did not indicate in what sense the solution Travonda gave was desirable, his expression of delight left no doubt that, in his view, this solution was special. As Voigt (1995) notes, such judgments serve an important function in supporting students' mathematical learning by making it possible for them to become aware of more conceptually advanced forms of mathematical activity while, at the same time, leaving it to them to decide whether to take up the intellectual challenge. Students can develop a sense of the teacher's expectations for their mathematical learning without feeling obliged to imitate solutions that might be beyond their current conceptual possibilities. In this case, several children took up the challenge of attempting to give solutions that they infer might also qualify as special. The episode continued as follows:

Chad: You, you can put the four [red chips] on that [right] side and you would make 10.
Teacher: *Yeah! I like that*.
Teacher: (To class) Chad says put these four (pointing to the red chips) over here (pointing to the blank spaces on the right frame) and that would make how many?
Class: Ten.
Teacher: Ten. *Okay, that's good*. Yeah?
Greg: Two plus 2 is four (pointing to the red chips) and 2 plus 2 is 4 (pointing to four green chips) and that's 8, and 2 more is 10.
Teacher: *Right. Do you understand what he said?* (The teacher repeats the solution for the class.)
John: You could do 7 plus 3 and then that would be 10.
Teacher: *I like that*.

Our observations indicated that all of the solutions that followed the teacher's enthusiastic response to Travonda's solution were novel for this class. For his part, the teacher continued to call attention to the solutions, indicating both that he wanted the other children to understand them and that he valued them. In the process, the sophistication both of individual children's thinking and of the mathematical discourse advanced. In Example 3, for instance, the solutions children offered became more sophisticated after the teacher indicated that he valued Travonda's solution. In this case, by sophisticated, we mean that the solutions went beyond counting by ones and involved constructing numerical relationships and developing alternative ways of combining elements of the two collections. John's comment, "*You could do* 7 plus 3 and then *that would be* 10," illustrates that children engaged in this type of extended activity. His language of "could do" and "that would be" suggests that, rather than reporting the way he initially solved the problem, he may be describing a relationship that he now realizes he could have used to solve the problem.

Influence of sociomathematical norms on mathematical argumentation and learning opportunities

We noted earlier that additional learning opportunities arise when children attempt to make sense of explanations given by others, to compare others' solutions to their own, and to make judgments about similarities and differences. Analysis of the children's activity shows that they constructed increasingly sophisticated concepts of ten, partitioned and recomposed two-digit numbers flexibly, and developed ways of talking about their mental activity using the standard language of tens and ones (Yackel, Cobb, & Wood, in press). Further, by explaining and justifying different solutions, the teacher and students established taken-as-shared meanings for tens and ones. In the

process, these became experientially real mathematical objects (Davis & Hersh, 1981) for almost all of the children in the class.

The negotiation of sociomathematical norms gives rise to learning opportunities for teachers as well as for students. One of the teacher's roles in an inquiry classroom is to facilitate mathematical discussions. At the same time, the teacher acts as a participant who can legitimize certain aspects of the children's mathematical activity and implicitly sanction others (Lampert, 1990; Voigt, 1985). Whole-class discussions are demanding situations for teachers because they have to try to make sense of the wide array of (different) solutions offered by the children (cf. Carpenter, Ansell, Franke, Fennema, & Weisbeck, 1993). Our observations consistently indicate that teachers capitalize on the learning opportunities that arise for them as they begin to listen to their students' explanations. The increasingly sophisticated way they select tasks and respond to children's solutions, shows their own developing understanding of the students' mathematical activity and conceptual development.

These learning opportunities for the teachers are directly influenced by the sociomathematical norms negotiated in the classrooms. In particular, children continue to give a variety of explanations when different solutions are emphasized and developmentally sophisticated solutions are legitimized. These inform the teachers about the students' conceptual possibilities and their current understandings. The latter, in turn, contribute to the teachers' evolving notions of what is sophisticated and efficient for the children. This further illustrates the reflexive relationship between the establishment of sociomathematical norms and the teacher's increasing understanding of mathematical difference, sophistication, and efficiency. For a more detailed discussion of teachers' learning in inquiry mathematics classrooms see Wood, Cobb, and Yackel (1991) and Yackel, Cobb, and Wood (in press).

The interactive constitution of what counts as an acceptable explanation and justification

We turn now to consider how the teacher and students in an inquiry mathematics classroom interactively constitute what counts as an acceptable explanation and justification and thus elaborate their taken-as-shared basis for communication. Viewed as a communicative act, explaining has as its purpose clarifying aspects of one's (mathematical) thinking that might not be apparent to others. Consequently, what is offered as an explanation is relative to the perceived expectations of others.

Our analysis of classroom data shows an evolution of students' understanding of what counts as an acceptable mathematical explanation and justification (Yackel, 1992). Initially, students' explanations may have a social rather than a mathematical basis. As their participation in inquiry mathematics instruction increases, they differentiate between various types of mathematical reasons. For example, they distinguish between explanations that describe

procedures and those that describe actions on experientially real mathematical objects. Finally, some students progress to being able to take explanations as objects of reflection. In the following discussion we illustrate these three aspects of students' understanding of explanation. In each case the focus of the discussion is on the interactive constitution of what constitutes acceptability.

A mathematical basis for explanations

A preliminary step in children's developing understanding of what constitutes an acceptable mathematical explanation is that they understand that the basis for their actions should be mathematical rather than status-based. Developing this preliminary understanding is not a trivial matter, especially since children are often socialized in school to rely on social cues for evaluation and on authority-based rationales. For example, in many classrooms it is appropriate for a child to infer that his answer is incorrect if the teacher questions it. In the classrooms that we have studied, one of the expectations is that children explain their solution methods to each other in small-group work and in whole-class discussions. However, most of the children were experiencing inquiry-based instruction for the first time and had little basis for knowing what types of rationales might be acceptable. In their prior experience of doing mathematics in school their teachers were typically the only members of the classroom community who gave explanations. They were therefore accustomed to relying on authority and status to develop rationales. For example, early in the school year one child attempted to resolve a dispute about an answer during small-group work by initiating a discussion about who had the best pencil and then about which of them was the smartest. This attempt to use status rather than a mathematical rationale to resolve the disagreement is consistent with the way many children interpret traditional mathematics instruction, as arbitrary procedures prescribed by their classroom authorities—the textbook and the teacher (Kamii, 1994; Voigt, 1992).

The following episode, which occurred early in the school year, demonstrates how a teacher can capitalize on situations that arise naturally in the classroom to make children's reasons an explicit topic of discussion.

Example 4: The teacher has posed a double tens-frame task using two red chips in the left tens-frame and 8 green chips in the right tens-frame.

Teacher: How many more green are there than red? How many more?
Donna: Six.
Teacher: There are six? All right. Six. Is that right class?
Students: Yes. No.
Donna: Oh, seven.
Student: Oh, I know.
Teacher: Seven.

Donna: Eight. Eight.
Student: I know. I know.
Teacher: (To Donna.) There are eight more green than there are red?
Student: No.
Student: Oh, Mr. K., I know.
Teacher: Think about it Donna. How many more green circles are there than red? Daria?
Daria: Six.
Teacher: How many?
Daria: Six.
Teacher: Is that right class? Do we agree with that?
Students: No. Yes.
Teacher: I heard some nos.
Many students begin talking at once.
Teacher: Listen. Listen.
Donna: (Protesting to the teacher) I said the six, but you said, "No."

In response to Donna's explicit acknowledgment that she changed her answers on the basis of her interpretation of the social situation rather than on mathematical reasoning, the teacher invents a scenario to clarify his expectations for this class.

Teacher: Wait, listen, listen. What did Mr. K.—what have I always taught you? (To Donna) What's your name?
Donna: My name is Donna Walters.
Teacher: What's your name?
Donna: My name is Donna Walters.
Teacher: If I were to ask you, "What's your name?" again, would you tell me your name is Mary?
Donna: No.
Teacher: Why wouldn't you?
Donna: Because my name is not Mary.
Teacher: And you know your name is—. . . . If you're not for sure you might have said your name is Mary. But you said Donna every time I asked you because what? You what? You know your name is what?
Donna: Donna.
Teacher: Donna. I can't make you say your name is Mary. So you should have said, "Mr. K. Six. And I can prove it to you." I've tried to teach you that.

Interventions of this type are powerful because they become paradigm cases that students can refer to. In general, such interventions are successful in establishing the expectation that rationales should be mathematical.

Explanations as descriptions of actions on experientially real mathematical objects

A more complex issue than establishing that mathematical reasons should form the basis for explanations, is *which types* of mathematical reasons might be acceptable. Here again, reflexivity is a key notion that guides our attempt to make sense of the classroom. We argue that what constitutes an acceptable mathematical reason is interactively constituted by the students and the teacher in the course of classroom activity. In the classroom studied, the children contributed to establishing an inquiry mathematics tradition by generating their own personally meaningful ways of solving problems instead of following procedural instructions. Further, their explanations increasingly involved describing actions on what to them were mathematical objects. In this sense, their explanations were conceptual rather than calculational (Thompson, Philipp, Thompson, & Boyd, 1994). In addition, children took seriously their obligation to try to make sense of the explanations of others. As a consequence, explanations were frequently challenged if they could be interpreted as relying on procedural instructions or if they used language that did not carry the significance of actions on taken-as-shared mathematical objects, which were experientially real for the students. These challenges in turn gave rise to situations for the teacher and students to negotiate what was acceptable as a mathematical explanation. The following illustrative episode, which occurred 2 months after the beginning of the school year, clarifies how the sociomathematical norm of what is acceptable as a mathematical explanation, is interactively constituted.

Example 5: The episode begins as Travonda is explaining her solution to the following problem.

> Roberto had 12 pennies. After his grandmother gave him some more, he had 25 pennies. How many pennies did Roberto's grandmother give him?

At Travonda's direction, the teacher writes

$$\begin{array}{r} 12 \\ \underline{+13} \end{array}$$

on the overhead projector. Thus far, her explanation involves specifying the details of how to write the problem using conventional vertical format. She continues.

Travonda: I said, one plus one is two, and 3 plus 2 is 5.
Teacher: All right, she said . . .
Rick: I know what she was talking about.
Teacher: Three plus 2 is 5, and one plus one is two.

Travonda's explanation can be interpreted as only procedural in nature. She has not made explicit reference to the value of the quantities the numerals signify nor clarified that the results should be interpreted as 25. Furthermore, in repeating her solution, the teacher modifies it to make it conform even more closely to the standard algorithm by proceeding from right to left. Several children simultaneously challenge the explanation.

Jameel:	(Jumping from his seat and pointing to the screen.) Mr. K. That's 20. That's 20.
Rick:	(Simultaneously) Un-uh. That's 25.
Several students:	That's 25. That's 25. He's talking about that.
Jameel:	Ten. Ten. That's taking a 10 right here . . . (walking up to the overhead screen and pointing to the numbers as he talks). This 10 and 10 (pointing to the ones in the tens column). That's 20 (pointing to the 2 in the 10s column).
Teacher:	Right.
Jameel:	And this is 5 more and it's 25.
Teacher:	That's right. It's 25.

Both Rick's challenge that the answer should be expressed as 25, rather than as two single digits and Jameel's challenge that the ones signify 10s and the two signifies 20 contribute to establishing the sociomathematical norm that explanations must describe actions on mathematical objects. Further, by acknowledging the challenges and accepting Jameel's clarification the teacher legitimized the ongoing negotiation of what is acceptable as an explanation in this classroom.

As a communicative act, explanation assumes a taken-as-shared stance (Rommetveit, 1985). Consequently, what constitutes an acceptable explanation is constrained by what the speaker and the listeners take as shared. But, as the above example shows, what is taken as shared is itself established during class discussions. Further, our analyses of discussions across the school year document that what is taken-as-shared mathematically evolves as the year progresses. Here, Jameel's clarification assumes that the conceptual acts of decomposing 12 into 10 and 2 and of decomposing 13 into 10 and 3 are shared by other students. Individual interviews conducted with all of the children in the class shortly before this episode occurred indicate that for a number of students this was not the case. Thus, although Jameel's explanation made it possible for him to orient his own understanding to Travonda's reported activity, it may have been inadequate for others.

Explanations as objects of reflection

When students begin to consider the adequacy of an explanation for others rather than simply for themselves, the explanation itself becomes the explicit

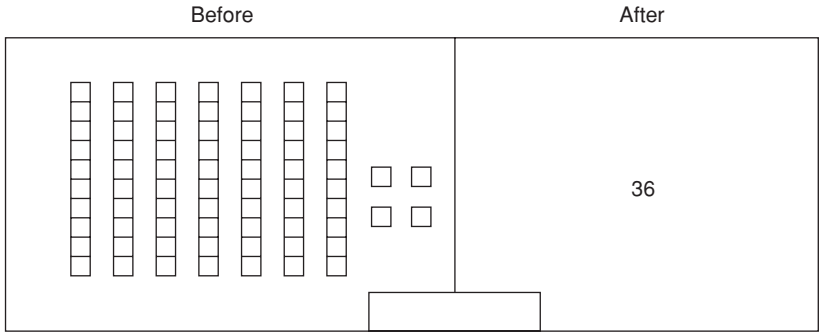


Figure 2 Problem task as shown on student activity page

object of discourse (Feldman, 1987). During classroom discussions, it is typically the teacher's responsibility to make implicit judgments about the extent to which students take something as shared and to facilitate communication by explicating the need for further explanation. As students' understanding of an acceptable explanation evolves, they too may assume this role. To do so, they must go beyond making sense of an explanation for themselves to making judgments about how other children might make sense of it. This involves a shift from participating in explanation to making the explanation itself an object of reflection. This shift in students' thinking is analogous to the shift between process and object that Sfard (1991) describes for mathematical conceptions. In the same way that being able to see a mathematical entity as an object as well as a process indicates a deeper understanding of the mathematical entity, taking an explanation as an object of reflection indicates a deeper understanding of what constitutes explanation.

The following example clarifies the shift in thinking that accompanies focusing on the explanation itself as an object. The episode occurred close to the end of the school year.

Example 6: Daria and Donna use centicubes on the overhead projector to explain their solution to the problem shown in Figure 2. The task is to figure out how much to add to or subtract from what is shown "before" to get what is shown "after." The girls had arrived at 38 as an answer during small-group work. To describe their solution to the class, they first place 74 centicubes on the overhead projector, using seven strips of ten (strips) and four individual cubes (squares).

Daria: We took this 40 off (points to four strips which the teacher then removes). That left 34. Oh, (to the teacher) put a 10 back. (The teacher replaces one of the strips.) 35, 36 (pointing to two of the cubes in the additional strip).

In our experience, purely conceptual solutions to tasks of this type require part-whole reasoning with tens and ones. This appears to be beyond the current conceptual capabilities of many second graders, and they needed to use manipulative or visual materials both to solve the tasks and to understand others' explanations. However, the strip (of 10 ones) the girls pointed to when they said, "35, 36" appeared as a single object on the overhead screen. Only those children who were looking directly at the materials laid on the overhead projector could see the 10 ones that composed the strip. The visual material available to the girls giving the explanation and to the children listening to the explanation, except for those children sitting immediately next to the overhead projector, was not the same. This subtle, but significant, point is indicated by Jameel's question.

Jameel: How—Wait, I got a question.
Teacher: Wait a minute, count that—
Jameel: Hey, Mr. K. If—How could she know, if you show two—How could the other person see if she does like when she said 44, 45, 46? How could she know it was two strips, I mean how could they know it was two squares like that? (Jameel appears to misspeak when he says 44, 45, 46 instead of 34, 35, 36.)
Toni: 'Cause they can see it.
Rick: No, we can't. We can't see it.

Jameel's question initiates a shift in the discussion from the solution of the problem to the adequacy and clarity of the explanation. At first glance, it may seem that his challenge is simply about the use of the manipulative materials. However, Toni's and Rick's responses and the subsequent discussion clarify that the issue is the coordination of tens and ones. Toni's reaction is interesting, given what we know about her conceptual possibilities. She is one of the children who would need to have manipulative or visual materials to solve the problem. However, she, like Jameel, was sitting immediately next to the overhead projector, and she looked at what Daria was actually pointing to rather than at what was visible on the overhead screen. Rick, however, is one of the children who would be able to solve the problem without using manipulatives. His "No, we can't. We can't see it," indicates that he shares Jameel's understanding that Daria's explanation has not clarified that the strip can be thought of as 10 ones.

The episode continues when the girls ask if there are any other questions. Jameel insists that the explanation requires elaboration, and the girls explain their solution again. Now, Daria actually removes 38 cubes in an attempt to demonstrate their solution. She removes three strips and the four individual cubes and breaks four additional cubes off of one of the remaining strips, leaving six connected cubes.

Students: Take those (strip of six) apart.
Teacher: Take those apart.

The girls break the six connected cubes apart, making it possible for all of the children to see them individually and therefore to count them. Finally, Daria counts to verify that there are 36, pointing as she counts, "10, 20, 30, 31, 32, 33, 34, 35, 36." This final explanation provides the explication that Jameel called for.

The preceding episode is significant because it shows that at least some of the children went beyond trying to make sense of an explanation for themselves and considered the extent to which it might be comprehensible to other members of the class. Jameel's criticism of the explanation was not that it didn't make sense to him. Rather, it was that those who could not see the 10 ones in the 10-strip might not be able to make sense of it. Jameel's question shifted the focus of the discussion from the solution of the problem to the adequacy of the explanation. In doing so, he made the explanation itself an object of reflection for others in the class as well as for himself.

Intellectual autonomy

The development of intellectual and social autonomy is a major goal in the current educational reform movement, more generally, and in the reform movement in mathematics education, in particular (National Council of Teachers of Mathematics, 1989). In this regard, the reform is in agreement with Piaget (1948/1973) that the main purpose of education is autonomy. Prior analysis shows that one of the benefits of establishing the social norms implicit in the inquiry approach to mathematics instruction is that they foster children's development of *social autonomy* (Cobb, et al., 1991; Cobb, Yackel, & Wood, 1989; Kamii, 1985; Nicholls, Cobb, Wood, Yackel, & Patashnick, 1990). However, it is the analysis of *sociomathematical norms* implicit in the inquiry mathematics tradition that clarifies the process by which teachers foster the development of intellectual autonomy.

In this account, the conception of autonomy as a context-free characteristic of the individual is rejected. Instead, autonomy is defined with respect to students' participation in the practices of the classroom community. In particular, students who are intellectually autonomous in mathematics are aware of, and draw on, their own intellectual capabilities when making mathematical decisions and judgments as they participate in these practices (Kamii, 1985). These students can be contrasted with those who are intellectually heteronomous and who rely on the pronouncements of an authority to know how to act appropriately. The link between the growth of intellectual autonomy and the development of an inquiry mathematics tradition becomes apparent when we note that, in such a classroom, the teacher guides the development of a community of validators and thus encourages the devolution

of responsibility. However, students can take over some of the traditional teacher's responsibilities only to the extent that they have constructed personal ways of judging that enable them to know in action both when it is appropriate to make a mathematical contribution and what constitutes an acceptable mathematical contribution. This requires, among other things, that students can judge what counts as a different solution, an insightful solution, an efficient solution, and an acceptable explanation. But, as we have attempted to illustrate throughout this paper, these are the types of judgments that the teacher and students negotiate when establishing sociomathematical norms that characterize an inquiry mathematics tradition. In the process, students construct specifically mathematical beliefs and values that help form their judgments. For instance, Jameel's challenge that "one and one is two" signifies "ten and ten is twenty" illustrates that children are capable of making judgments about what is appropriate mathematically. Further, Jameel's challenge indicates that he had developed the belief that mathematical explanations should describe actions on experientially real mathematical objects. Examples such as this show that it is precisely because children can make personal judgments of this kind on the basis of their mathematical beliefs and values that they can participate as increasingly autonomous members of an inquiry mathematics community.

Significance

The notion of sociomathematical norms that we have advanced in this paper is important because it sets forth a way of analyzing and talking about the *mathematical* aspects of teachers' and students' activity in the mathematics classroom. This is a significant extension of prior work on general classroom social norms in that it clarifies aspects of teachers' and students' activity that sustain a classroom atmosphere conducive to problem solving and inquiry. These sociomathematical norms are intrinsic aspects of the classroom's mathematical microculture. Nevertheless, although they are specific to mathematics, they cut across areas of mathematical content by dealing with mathematical qualities of solutions, such as their similarities and differences, sophistication, and efficiency. Additionally, they encompass ways of judging what counts as an acceptable mathematical explanation.

We have also attempted to demonstrate that these norms are not predetermined criteria introduced into the classroom from the outside. Instead, these normative understandings are continually regenerated and modified by the students and the teacher through their ongoing interactions. As teachers gain experience with an inquiry approach to mathematics instruction they may have some clear ideas in advance of norms that they might wish to foster. Even in such cases these norms are, of necessity, interactively constituted by each classroom community. Consequently, the sociomathematical norms that are constituted might differ substantially from one classroom to

another. For purposes of this paper, we have discussed the development of sociomathematical norms in classrooms that generally follow an inquiry form of instruction. As we have shown, in the process of negotiating sociomathematical norms, students in these classrooms actively constructed personal beliefs and values that enabled them to be increasingly autonomous in mathematics.

The notion of sociomathematical norms is also important for clarifying the teacher's role as a representative of the mathematical community. The question of the teacher's role in classrooms that attempt to develop a practice consistent with the current reform emphasis on problem solving and inquiry is one of current debate (Clement, 1991). Many teachers assume that they are expected to assume a passive role (P. Human, personal communication, August 1994). However, we question this position. As we have stated previously,

> The conclusion that teachers should not attempt to influence students' constructive efforts seems indefensible, given our contention that mathematics can be viewed as a social practice or a community project. From our perspective, the suggestion that students can be left to their own devices to construct the mathematical ways of knowing compatible with those of wider society is a contradiction in terms. (Cobb, Yackel, & Wood, 1992, pp. 27–28)

In this paper we have attempted to clarify one critical aspect of the teacher's role in influencing the mathematical aspects of the knowledge children construct. In this regard, the ideas set forth in this paper are potentially useful in preservice and inservice teacher education. For example, in a recent project classroom teaching experiment, the notion of sociomathematical norms influenced discussions between the researcher and the classroom teacher. In particular, the issue of what constitutes a mathematically efficient solution became an explicit focus in discussions with the teacher and in the classroom itself. In the process, the level of discourse and the individual children's learning advanced (Cobb, Boufi, McClain, & Whitenack, in press).

The analysis of sociomathematical norms indicates that the teacher plays a central role in establishing the mathematical quality of the classroom environment and in establishing norms for mathematical aspects of students' activity. It further highlights the significance of the teacher's own personal mathematical beliefs and values and their own mathematical knowledge and understanding. In this way, the critical and central role of the teacher as a representative of the mathematical community is underscored.

Acknowledgments

A previous version of this paper was presented at the 1993 annual meeting of the American Educational Research Association, Atlanta, GA.

Several notions central to this paper were elaborated in the course of discussions with Heinrich Bauersfeld, Gotz Krummheuer, and Jorg Voigt of the University of Bielefeld, Germany and Terry Wood of Purdue University.

The research reported in this paper was supported by the National Science Foundation under grant numbers RED-9353587, DMS-9057141 and MDR 885-0560, by the James S. McDonnell Foundation, and by the Spencer Foundation. All opinions expressed are solely those of the authors.

References

Bauersfeld, H. (1988). Interaction, construction, and knowledge: Alternative perspectives for mathematics education. In T. Cooney & D. Grouws (Eds.), *Effective mathematics teaching* (pp. 27–46). Reston, VA: National Council of Teachers of Mathematics/Erlbaum.

Bauersfeld, H. (1993, March). Teachers pre and in-service education for mathematics teaching. *Seminaire sur la Representation*, No. 78, CIRADE, Université du Québec à Montréal, Canada.

Bauersfeld, H., Krummheuer, G., & Voigt, J. (1988). Interactional theory of learning and teaching mathematics and related microethnographical studies. In H. G. Steiner & A. Vermandel (Eds.), *Foundations and methodology of the discipline of mathematics education* (pp. 174–188). Antwerp: Proceedings of the Theory of Mathematics Education Conference.

Blumer, H. (1969). *Symbolic interactionism*. Engelwood Cliffs, NJ: Prentice-Hall.

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education*, *24*, 427–440.

Clement, J. (1991). Constructivism in the classroom [Review of the book *Transforming children's mathematics education: International perspectives*]. *Journal for Research in Mathematics Education*, *22*, 422–428.

Cobb, P. (1990). Multiple Perspectives. In L. P. Steffe & T. Wood (Eds.), *Transforming children's mathematics education: International perspectives* (pp. 200–215). Hillsdale, NJ: Erlbaum.

Cobb, P., & Bauersfeld, H. (Eds.). (1995). *Emergence of mathematical meaning: Interaction in classroom cultures*. Hillsdale, NJ: Erlbaum.

Cobb, P., Boufi, A., McClain, K., & Whitenack, J. (in press). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education*.

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of classroom mathematics traditions: An interactional analysis. *American Educational Research Journal*, *29*, 573–604.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education*, *22*, 3–9.

Cobb, P., Yackel, E., & Wood, T. (1989). Young children's emotional acts while doing mathematical problem solving. In D. B. McLeod & V. M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective* (pp. 117–148). New York: Springer-Verlag.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education*, *23*, 2–33.

Davis, P. J., & Hersh, R. (1981). *The mathematical experience*. Boston: Houghton Mifflin.

Eisenhart, M. A. (1988). The ethnographic research tradition and mathematics education research. *Journal for Research in Mathematics Education*, *19*, 99–114.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 119–161). New York: Macmillan.

Feldman, C. F. (1987). Thought from language: The linguistic construction of cognitive representations. In J. Bruner & H. Haste (Eds.), *Making sense: The child's construction of the world* (pp. 131–162). London: Methuen.

Greeno, J. (1991). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education*, *22*, 170–218.

Kamii, C. (1985). *Young children reinvent arithmetic: Implications of Piaget's theory*. New York: Teachers College Press.

Kamii, C. (1994). *Young children continue to reinvent arithmetic—3rd grade: Implications of Piaget's theory*. New York: Teachers College Press.

Krummheuer, G. (1983). Das arbeitsinterim im mathematikunterricht [The working interim in mathematics classrooms]. In H. Bauersfeld (Ed.) *Lernen und lehren von mathematik. Analysen zum unterrichishandeln* (pp. 57–106). Köln, Germany: Aulis.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, *27*, 29–63.

Leiter, K. (1980). *A primer on ethnomethodology*. New York: Oxford University Press.

Mehan, H., & Wood, H. (1975). *The reality of ethnomethodology*. New York: John Wiley.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

Nicholls, J., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Dimensions of success in mathematics: Individual and classroom differences. *Journal for Research in Mathematics Education*, *21*, 109–122.

Piaget, J. (1973). *To understand is to invent*. New York: Grossman. (Original work published 1948)

Resnick, L. B. (1989). *Knowing, learning, and instruction*. Hillsdale, NJ: Erlbaum.

Richards, J. (1991). Mathematical discussions. In E. von Glasersfeld (Ed.), *Constructivism in mathematics education* (pp. 13–52). Dordrecht, The Netherlands: Kluwer.

Rommetveit, R. (1985). Language acquisition as increasing linguistic structuring of experience and symbolic behavior control. In J. V. Wertsch (Ed.), *Culture, communication, and cognition* (pp. 183–205). Cambridge: Cambridge University Press.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*, 1–36.

Thompson, A. G., Philipp, R. A., Thompson, P. W., & Boyd, B. (1994). Calculational and conceptual orientations in teaching mathematics. In D. Aichele & A. F. Coxford

(Eds.), *Professional development of teachers of mathematics* (pp. 79–92). Reston, VA: National Council of Teachers of Mathematics.

Voigt, J. (1985). Patterns and routines in classroom interaction. *Recherches en Didactique des Mathématiques*, *6*, 69–118.

Voigt, J. (1989). The social constitution of the mathematics province—A microethnographical study in classroom interaction. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, *11*(1 & 2), 27–34.

Voigt, J. (1992, August). *Negotiation of mathematical meaning in classroom practices: Social interaction and learning mathematics*. Paper presented at the Seventh International Congress on Mathematical Education, Quebec City.

Voigt, J. (1995). Thematic patterns of interaction and sociomathematical norms. In P. Cobb & H. Bauersfeld (Eds.), *Emergence of mathematical meaning: Interaction in classroom cultures* (pp. 163–201). Hillsdale, NJ: Erlbaum.

von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality* (pp. 17–40). New York: Norton.

Wood, T., Cobb, P., & Yackel, E. (1991). Change in teaching mathematics: A case study. *American Educational Research Journal*, *28*, 587–616.

Yackel, E. (1992, August). *The evolution of second grade children's understanding of what constitutes an explanation in a mathematics class*. Paper presented at the Seventh International Congress of Mathematics Education, Quebec City.

Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education*, *22*, 390–408.

Yackel, E., Cobb, P., & Wood, T. (in press). The interactive constitution of mathematical meaning in one second grade classroom: An illustrative example. *Journal of Mathematical Behavior*.

73

SEX DIFFERENCES IN MATHEMATICAL ABILITY

Fact or artifact?

C. P. Benbow and J. C. Stanley

A substantial sex difference in mathematical reasoning ability (score on the mathematics test of the Scholastic Aptitude Test) in favor of boys was found in a study of 9927 intellectually gifted junior high school students. Our data contradict the hypothesis that differential course-taking accounts for observed sex differences in mathematical ability, but support the hypothesis that these differences are somewhat increased by environmental influences.

Huge sex differences have been reported in mathematical aptitude and achievement (*1*). In junior high school, this sex difference is quite obvious: girls excel in computation, while boys excel on tasks requiring mathematical reasoning ability (*1*). Some investigators believe that differential course-taking gives rise to the apparently inferior mathematical reasoning ability of girls (*2*). One alternative, however, could be that less well-developed mathematical reasoning ability contributes to girls' taking fewer mathematics courses and achieving less than boys.

We now present extensive data collected by the Study of Mathematically Precocious Youth (SMPY) for the past 8 years to examine mathematical aptitude in approximately 10,000 males and females prior to the onset of differential course-taking. These data show that large sex differences in mathematical aptitude are observed in boys and girls with essentially identical formal educational experiences.

Six separate SMPY talent searches were conducted (*3*). In the first three searches. 7th and 8th graders, as well as accelerated 9th and 10th graders, were eligible: for the last three, only 7th graders and accelerated students of 7th grade age were eligible. In addition, in the 1976, 1978, and 1979 searches,

Source: *Science*, 1980, 210(2), 1262–1264.

the students had also to be in the upper 3 percent in mathematical ability as judged by a standardized achievement test, in 1972 in the upper 5 percent, and in 1973 and 1974 in the upper 2 percent. Thus, both male and female talent-search participants were selected by equal criteria for high mathematical ability before entering. Girls constituted 43 percent of the participants in these searches.

As part of each talent search the students took both parts of the College Board's Scholastic Aptitude Test (SAT)—the mathematics (SAT-M) and the verbal (SAT-V) tests (*4*). The SAT is designed for able juniors and seniors in high school, who are an average of 4 to 5 years older than the students in the talent searches. The mathematical section is particularly designed to measure mathematical reasoning ability (*5*). For this reason, scores on the SAT-M achieved by 7th and 8th graders provided an excellent opportunity to test the Fennema and Sherman differential course-taking hypothesis (*2*), since until then all students had received essentially identical formal instruction in mathematics (*6*). If their hypothesis is correct, little difference in mathematical aptitude should be seen between able boys and girls in our talent searches.

Results from the six talent searches are shown in Table 1. Most students scored high on both the SAT-M and SAT-V. On the SAT-V, the boys and girls performed about equally well (*7*). The overall performance of 7th grade students on SAT-V was at or above the average of arandom sample of high school students, whose mean score is 368 (*8*), or at about the 30th percentile of college-bound 12th graders. The 8th graders, regular and accelerated, scored at about the 50th percentile of college-bound seniors. This was a high level of performance.

A large sex difference in mathematical ability in favor of boys was observed in every talent search. The smallest mean difference in the six talent searches was 32 points in 1979 in favor of boys. The statistically significant *t*-tests of mean differences ranged from 2.5 to 11.6 (*9*). Thus, on the average, the boys scored about one-half of the females' standard deviation (S.D.) better than did the girls in each talent search, even though all students had been certified initially to be in the top 2nd, 3rd, or 5th percentiles in mathematical reasoning ability (depending on which search was entered).

One might suspect that the SMPY talent search selected for abler boys than girls. In all comparisons except for two (8th graders in 1972 and 1976), however, the girls performed better on SAT-M relative to female college-bound seniors than the boys did on SAT-M relative to male college-bound seniors. Furthermore, in all searches, the girls were equal verbally to the boys. Thus, even though the talent-search girls were at least as able compared to girls in general as the talent-search boys were compared to boys in general, the boys still averaged considerably higher on SAT-M than the girls did.

Moreover, the greatest disparity between the girls and boys is in the upper ranges of mathematical reasoning ability. Differences between the top-scoring

Table 1 Performance of students in the Study of Mathematically Precocious Youth in each talent search ($N = 9927$)

		Number		*SAT-V score** ($\bar{X} \pm$ S.D.)		*SAT-M scores*†				*Percentage scoring above 600 on SAT-M*	
						$\bar{X} \pm$ S.D.		*Highest score*			
Test date	*Grade*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
March 1972	7	90	77			460 ± 104	423 ± 75	740	590	7.8	0
	8†	133	96			528 ± 105	458 ± 88	790	600	27.1	0
January 1973	7	135	88	385 ± 71	374 ± 74	495 ± 85	440 ± 66	800	620	8.1	1.1
	8†	286	158	431 ± 89	442 ± 83	551 ± 85	511 ± 63	800	650	22.7	8.2
January 1974	7	372	222			473 ± 85	440 ± 68	760	630	6.5	1.8
	8†	556	369			540 ± 82	503 ± 72	750	700	??.6	7.9
December 1976	7	495	356	370 ± 73	368 ± 70	455 ± 84	421 ± 64	780	610	5.5	0.6
	8‡	12	10	487 ± 129	390 ± 61	598 ± 126	482 ± 83	750	600	58.3	0
January 1978	7 and 8‡	1549	1249	375 ± 80	372 ± 78	448 ± 87	413 ± 71	790	760	5.3	0.8
January 1979	7 and 8‡	2046	1628	370 ± 76	370 ± 77	436 ± 87	404 ± 77	790	760	3.2	0.9

* Mean score for a random sample of high school juniors and seniors was 368 for males and females (*8*).

† Mean for juniors and seniors: males, 416; females, 390 (*8*).

‡ These rare 8th graders were accelerated at least 1 year in school grade placement.

boys and girls have been as large as 190 points (1972 8th graders) and as low as 30 points (1978 and 1979). When one looks further at students who scored above 600 on SAT-M, Table 1 shows a great difference in the percentage of boys and girls. To take the extreme (not including the 1976 8th graders), among the 1972 8th graders, 27.1 percent of the boys scored higher than 600, whereas not one of the girls did. Over all talent searches, boys outnumbered girls more than 2 to 1 (1817 boys versus 675 girls) in SAT-M scores over 500. In not one of the six talent searches was the top SAT-M score earned by a girl. It is clear that much of the sex difference on SAT-M can be accounted for by a lack of high-scoring girls.

A few highly mathematically able girls have been found, particularly in the latest two talent searches. The latter talent searches, however, were by far the largest, making it more likely that we could identify females of high mathematical ability. Alternatively, even if highly able girls have felt more confident to enter the mathematics talent search in recent years, our general conclusions would not be altered unless all of the girls with the highest ability had stayed away for more than 5 years. We consider that unlikely, In this context, three-fourths as many girls have participated as boys each year; the relative percentages have not varied over the years.

It is notable that we observed sizable sex differences in mathematical reasoning ability in 7th grade students. Until that grade, boys and girls have presumably had essentially the same amount of formal training in mathematics. This assumption is supported by the fact that in the 1976 talent search no substantial sex differences were found in either participation in special mathematics programs or in mathematical learning processes (*6*). Thus, the sex difference in mathematical reasoning ability we found was observed before girls and boys started to differ significantly in the number and types of mathematics courses taken. It is therefore obvious that differential course-taking in mathematics cannot alone explain the sex difference we observed in mathematical reasoning ability, although other environmental explanations have not been ruled out.

The sex difference in favor of boys found at the time of the talent search was sustained and even increased through the high school years. In a follow-up survey of talent-search participants who had graduated from high school in 1977 (*10*), the 40-point mean difference on SAT-M in favor of boys at the time of that group's talent search had increased to a 50-point mean difference at the time of high school graduation. This subsequent increase is consistent with the hypothesis that differential course-taking can affect mathematical ability (*2*). The increase was rather small, however. Our data also show a sex difference in the number of mathematics courses taken in favor of boys but not a large one. The difference stemmed mainly from the fact that approximately 35 percent fewer girls than boys took calculus in high school (*10*). An equal proportion of girls and boys took mathematics in the 11th grade (83 percent), however, which is actually the last grade completed before

taking the SAT in high school. It, therefore, cannot be argued that these boys received substantially more formal practice in mathematics and therefore scored better. Instead, it is more likely that mathematical reasoning ability influences subsequent differential course-taking in mathematics. There were also no significant sex differences in the grades earned in the various mathematics courses (*10*).

A possible criticism of our results is that only selected mathematically able, highly motivated students were tested. Are the SMPY results indicative of the general population? Lowering qualifications for the talent search did not result in more high-scoring individuals (except in 1972, which was a small and not widely known search), suggesting that the same results in the high range would be observed even if a broader population were tested. In addition, most of the concern about the lack of participation of females in mathematics expressed by Ernest (*11*) and others has been about intellectually able girls, rather than those of average or below average intellectual ability.

To what extent do girls with high mathematical reasoning ability opt out of the SMPY talent searches? More boys than girls (57 percent versus 43 percent) enter the talent search each year. For this to change our conclusions, however, it would be necessary to postulate that the most highly talented girls were the least likely to enter each search. On both empirical and logical grounds this seems improbable.

It is hard to dissect out the influences of societal expectations and attitudes on mathematical reasoning ability. For example, rated liking of mathematics and rated importance of mathematics in future careers had no substantial relationship with SAT-M scores (*6*). Our results suggest that these environmental influences are more significant for achievement in mathematics than for mathematical aptitude.

We favor the hypothesis that sex differences in achievement in and attitude toward mathematics result from superior male mathematical ability, which may in turn be related to greater male ability in spatial tasks (*12*). This male superiority is probably an expression of a combination of both endogenous and exogenous variables. We recognize, however, that our data are consistent with numerous alternative hypotheses. Nonetheless, the hypothesis of differential course-taking was not supported. It also seems likely that putting one's faith in boy-versus-girl socialization processes as the only permissible explanation of the sex difference in mathematics is premature.

References and notes

1 E. Fennema, *J. Res. Math. Educ*. 5, 126 (1974). "National assessment for educational progress." *NAEP Neral.* 8 (No. 5). Insert (1975); L. Fox. in *Intellectual Talent Research and Development*, D. Keating. Ed. (Johns Hopkins Univ. Press, Baltimore, 1976), p. 183.

2 For example, E. Fennema and J. Sherman, *Am. Educ. Res. J.* **14**, 51 (1977).
3 W. George and C. Solano, in *Intellectual Talent: Research and Development*, D. Keating, Ed. (Johns Hopkins Univ. Press, Baltimore, 1976), p. 55.
4 The SAT-V was not administered in 1972 and 1974, and the Test of Standard Written English was required in 1978 and 1979.
5 W. Angoff, Ed., *The College Board Admissions Testing Program* (College Entrance Examination Board, Princeton, N.J., 1971), p. 15.
6 C. Benbow and J. Stanley, manuscript in preparation.
7 This was not true for the accelerated 8th graders in 1976. The N for the latter comparison is only 22.
8 College Entrance Examination Board, *Guide to the Admissions Testing Service* (Educational Testing Service, Princeton, N.J., 1978), p. 15.
9 The t-tests and P values for 7th and 8th graders, respectively, in the six talent searches were 2.6, $P < .01$; 5.3, $P < .001$; 5.1, $P < .001$; 5.2. $P < .001$; 4.9, $P < .001$; 7.1, $P < .001$; 6.6. $P < .001$; 2.5, $P < .05$; 11.6, $P < .001$; and 11.5, $P < .001$.
10 C. Benbow and J. Stanley, in preparation.
11 J. Ernest, *Am. Math. Mon.* **83**, 595 (1976).
12 I. MacFarlane-Smith. *Spatial Ability* (Univ. of London Press, London, 1964), J. Sherman, *Psychol. Rey*. **74**, 290 (1967).
13 We thank R. Benbow, C. Breaux, and L. Fox for their comments and help in preparing this manuscript. Supported in part by grants from the Spencer Foundation and the Educational Foundation of America.

Part XIV

SCIENCE, SOCIAL SCIENCE

74

THE ACQUISITION OF CONCEPTUAL KNOWLEDGE IN SCIENCE BY PRIMARY SCHOOL CHILDREN

Group interaction and the understanding of motion down an incline

C. Howe, A. Tolmie and C. Rodgers

It is widely accepted that primary school children will approach science with strong 'alternative conceptions' about the variables at play which, unless directly challenged, will circumscribe learning. Extensive discussion concerning the form the challenges should take has led to the conclusion that learning will be maximized if children explore their conceptions while working with peers whose alternative conceptions are different. At present, however, there is little research to support this, and the small amount that does exist says little about the process by which learning is effected. The current study attempted to redress this in the context of motion down an incline. Individual pre-tests were administered to 113 8- to 12-year-old children to establish their alternative conceptions. On the basis of their pre-test responses, and in order to establish adequate controls, the children were put into groups of four according to whether their conceptions were different or similar. The children worked in their groups on tasks designed to elicit the exploration of alternative conceptions, and were subsequently post-tested. The pattern of pre- to post-test change gave some support to the notion that learning is maximized when alternative conceptions differ. However, it gave few grounds for thinking that learning involves the internalization of conceptions that the groups jointly construct. Rather, it suggested a process of private conflict resolution, for which the catalyst was discussion held during the groups but continuing long after their completion.

Source: *British Journal of Developmental Psychology*, 1992, 10(2), 113–130.

In the past, research into the acquisition of conceptual knowledge in science by primary school children was seldom contemplated. Educationalists were mainly concerned with the secondary age range, and when conceptual knowledge was studied by psychologists, the logical and social domains were the central focus. Recently, however, there has been a change, and it is not hard to see why. After a decade of debate, the National Curriculum (Department of Education & Science, 1989) has stipulated that 'the knowledge and understanding of science' be taught from the first years of schooling. Thus, primary teachers have been charged with finding appropriate methods, and there was a widely held impression that research with the secondary age range would have little to tell them. It was known that prominent reviews like McDermott (1984) were documenting widespread failure to get conceptual knowledge across, with students often entering university with only the vaguest grasp of fundamental notions. Hence, it was felt that little could be gleaned from secondary practice apart from a need for different methods. Further research would be needed for positive suggestions, and this is what produced the momentum for the recent research.

Much of the research has been inspired by the view that children will come to primary school science not as 'blank slates', but with strong alternative conceptions about the issues at stake. Thus, the problem, as articulated by Hewson & Hewson (1983), will not so much be to write in the received wisdom as to change ideas in the appropriate direction. The proposed solutions vary depending on the precise nature of the conceptual knowledge under scrutiny. However, when it is understanding of the relevant variables, a popular approach has been one that involves children in making their alternative conceptions explicit and subjecting these to empirical test. It is a solution that has already been incorporated into published teaching materials and, in that sense, is readily translatable to classroom usage. However, resource limitations mean that the materials will almost certainly be presented to children in groups, and this has been a consideration in the recent research. It has been hypothesized that given group presentation, the composition of the groups is by no means irrelevant. On the contrary, if the groups comprise children whose alternative conceptions differ, their interaction will be such as to maximize learning.

The motivation for the hypothesis is partly the theorizing of Piaget. This is because Piaget (e.g. Piaget, 1972) clearly saw alternative conceptions about the variables which science makes relevant as the kind of ideas that advance by equilibration. As is well known, Piaget (1985) not only believed equilibration to be activated when opposing but incomplete conceptions give rise to internally experienced conflict. He also saw interaction over conceptions between children with differing and incomplete perspectives as a context where such conflict should arise (Piaget, 1932). Piaget, however, never tested his ideas empirically, and it was left to Doise and his associates to take things further. Their studies in the logical and spatial domains (now summarized in

Doise & Mugny, 1984) and the follow-ups by, for example, Ames & Murray (1982), Berkowitz, Gibbs & Broughton (1980), Damon & Killen (1982) and Weinstein & Bearison (1985) have added further weight to the hypothesis in science.

Glancing at the research which the hypothesis has stimulated, it might appear as if the advantages of groups where alternative conceptions differ has already been shown. In studies which required children to explicate, discuss and test their conceptions about the variables at play, Champagne, Gunstone & Klopfer (1983), Forman & Cazden (1985), Forman & Kraker (1985), Nussbaum & Novick (1981) and Osborne & Freyberg (1985), obtained results which are seemingly positive, giving the impression of ample support. On closer scrutiny, however, there are a number of problems. In some cases, the guarantees that the alternative conceptions differed were far from convincing. In others, the interaction was subject to the interpolation of 'expert' ideas (usually, though not always, from teachers), and this may have been producing the effects rather than the exchanges between the children. When these difficulties were avoided, the studies rarely had control groups of children with similar conceptions to differentiate the effects of group composition from the effects of empirical testing, and they seldom considered whether the beneficial outcomes survived over time.

An attempt to avoid such problems while researching the basic issue has, however, been reported by Howe, Tolmie & Rodgers (1990). It involved two studies, both concerned with knowledge of the variables relevant to flotation. In both, children aged 8 to 12 were pre-tested to establish their alternative conceptions, and grouped such that these conceptions were either different or similar. The children worked in their groups on tasks designed to elicit the discussion and empirical appraisal of alternative conceptions, and a few weeks later they were post-tested. In both studies, the children who worked in groups where alternative conceptions differed showed significantly greater progress from pre- to post-test, providing support for the hypothesis in the context of flotation. Flotation is, however, only one of the topics which, under the National Curriculum, primary school children will have to master, and it is not clear that similar outcomes would be obtained elsewhere.

The generalizability to other contexts is particularly unclear given the nature of the differing groups in Howe *et al.*'s (1990) study. Consistent with earlier research, Howe *et al.* found their subjects differing firstly in which of the countless possible irrelevant variables they habitually invoked, and secondly in whether the irrelevant variables were supplemented with relevant ones. Consequently, it was this kind of difference that their differing groups reflected. Reviews like Driver, Guesne & Tiberghien (1985) and Piaget (1930, 1974) make it clear that 8- to 12-year-olds differ in a similar fashion with other topics. Nevertheless, there are exceptions and studies by Ferretti, Butterfield, Cahn & Kerkman (1985), Inhelder & Piaget (1958) and Stead & Osborne (1981) suggest that motion down an incline is one of these. According

to these studies, 8- to 12-year-olds differ little over irrelevancies, for object weight is the only irrelevant variable commonly referred to. Equally, they differ little over supplementation with relevancies, for all the relevant variables are habitually acknowledged. Where 8- to 12-year-olds do differ is over how they deploy the relevant variables. Some are confused about how the variables operate, thinking for example that a steep angle inhibits motion. Others avoid confusion, but cannot coordinate the variables into an integrated model. A third group (usually at the upper end of the age range) can coordinate, and only have the details left to derive.

Given the contrasts with flotation, it would be helpful to see whether grouping such that alternative conceptions differ has beneficial effects with motion down an incline, and this was one aim of the study reported in this paper. In this sense, the study attempted to replicate Howe *et al.* with a contrasting topic. Replication was not, however, the study's only aim for, assuming the effects of group composition to be mediated through interaction, it sought also to supplement Howe *et al.* on the process by which this occurs. Doise & Mugny (1984) would seem to anticipate a process whereby group-generated conflict stimulates the joint construction of a superior conception which is then individually internalized. However, neither of Howe *et al.*'s studies supported this. In one, children whose group performance was worse than their pre-test were as likely to advance from pre- to post-test as children whose group performance was better. In the other, pre- to post-test change was positively correlated with group performance, but children whose performance was in opposition to other group members were as likely to advance as children whose performance was joint. Asking what factors other than internalization are precipitated by interaction, Howe *et al.* noted that it could be the continuation of changes made privately within the group. Alternatively (or in addition), it could be the adoption of changes made after the group's completion. If the latter, it could be with reference solely to the interaction or it could involve information solicited later. Recognizing these possibilities, the study aimed to shed light on each.

Method

Design

Pre-tests were administered to 113 8- to 12-year-old children to assess their alternative conceptions about the variables relevant to motion down an incline. Using the pre-tests, 84 of the children were assigned to groups of four such that alternative conceptions were either different or similar. Some six weeks later, the children worked in their groups on a task designed to elicit the explication, discussion and empirical testing of alternative conceptions. After completing the task, 25 per cent were given immediate post-tests to estimate private change within the group. All 84 were given delayed

post-tests around four weeks later. These post-tests were designed not simply to assess conceptual change, but also to investigate the solicitation of information after the task.

Subjects

The children were all pupils at the same inner Glasgow primary school. They were randomly selected from four age bands: primary four (8 to 9 years), primary five (9 to 10 years), primary six (10 to 11 years) and primary seven (11 to 12 years). Roughly equal numbers were chosen from each age band. Out of the total sample, 38 per cent of the children were of Asian origin, primarily from the Indian subcontinent.

Apparatus

The pre-tests, group task and post-tests all used four toy vehicles. These were two lorries, identical in appearance but different in weight; and two cars, identical in both appearance and weight, the latter being between that of the lorries. Thus, there were vehicles of what will be called 'light', 'middle' and 'heavy' object weights.

The vehicles were used with four parallel slopes which were supported by a vertical frame. The slopes were 1 m in length and 6 cm in width, and rested on pegs inserted into the frame. These pegs could be positioned 8.7, 19.3 and 42.4 cm from the ground to incline the slopes at 'low', 'middle' and 'steep' angles. Three gates were located at 10, 59 and 80 cm from the top of each slope. These gates could be open or closed to produce 'high', 'middle' and 'low' starting positions. Two of the slopes were covered with identical surfaces, the third with a lower friction surface and the fourth with a higher. Thus, the apparatus also allowed for 'high', 'middle' and 'low' surface friction.

The slopes terminated, via short flexible extensions, on a mat which was divided into 'near', 'middle' and 'far' areas.

Materials

(a) Pre- and post-test interview schedules

The apparatus permitted the manipulation of the three variables which are relevant to motion down an incline (angle, starting position and surface friction) and the one which though irrelevant was known from the literature cited earlier to be favoured by children (object weight). Within the pre- and post-tests, manipulation of the variables formed the basis for interview schedules which examined understanding of their independent and coordinated operation. There were two schedules, one for the pre-test and immediate post-test and one for the delayed post-test. They differed in content but

both provided six opportunities to respond on each of the relevant variables and eight to respond on the irrelevant. They did this through six three-stage and five single-stage items.

The three-stage items presupposed that a middle friction slope had been set up with the middle angle and middle starting position, and that one of the cars (i.e. a middle-weight vehicle) had been allowed to roll down and come to rest in the middle area. This constituted the 'standard display'. The first stage of each item presupposed that another slope had been adjusted, such that it (or the vehicle to be rolled down it) differed from the standard on one of the variables but was identical on the others (e.g. the middle-friction slope with the middle angle and the middle-weight vehicle but the low starting position). The second and third stages presupposed further adjustments, such that there were differences from the standard on two of the variables (e.g. the middle-friction slope with the middle weight but the steep angle and the low starting position) and then on three (e.g. the middle-friction slope with the heavy-weight vehicle, the steep angle and the low starting position). At each stage, subjects were asked to predict whether the vehicle would travel to the same area as in the standard, the near area or the far, and to explain their answers. It was assumed that subjects would reveal their alternative conceptions about the variables through the explanations they gave. Object weight was manipulated on all six items, with two of the manipulations at each of the stages. The other variables were manipulated on four items, with at least one of the manipulations at each of the stages.

The single-stage items were presented between the three-stage ones. They described real-world instances like two skateboarders, one on a gentle, icy slope and the other on a steep, ice-free slope, and two lorries, both free-wheeling on the same slope but one empty and the other loaded with bricks. Here, subjects were asked to predict which (if any) would travel furthest from the foot of the slope, and to explain their answers. Again, it was assumed that alternative conceptions would be revealed through the explanations. Two of the items manipulated object weight. The other three manipulated two of angle, starting position and surface friction.

The delayed post-test schedule concluded with three additional items. These required subjects to say whether they could find out more about rolling down slopes from, respectively, books, other people and direct observation, and if so whether they had tried to do this after the group task.

(b) Group task instruction book

Using a method shown by Howe *et al.* (1990) to be particularly successful at eliciting discussion, the group task comprised an individual phase followed by a collaborative one. For the individual phase, the apparatus was to be used with sets of six cards. The cards presupposed the standard display. They each asked subjects to tick whether the same area, the near area or

the far area would be reached after a change from the standard on one variable.

For the collaborative phase, the apparatus and cards were to be used with a book which provided detailed instructions on how to proceed. An extract is reproduced in Appendix I. The book presented six three-stage and five single-stage items. These items were similar in form but different in content from those appearing in the pre- and post-tests. For each of the three-stage items, the book invited subjects to create a display which differed from the standard on one variable. For guidance, the book provided an illustration of the display to be created, omitting the 'static' elements (i.e. the slopes that were not to be adjusted and the areas on the floor) to avoid clutter. Then the book requested subjects to compare the responses on the relevant cards and, when these responses differed, to come to an agreement. Once subjects had agreed a prediction, they were invited to test it, and agree an explanation when the outcome was different from what they expected. After doing this, they were asked to agree and test predictions and agree explanations given changes from the standard on two and then three variables, again following text which provided illustrations of the displays to be created. For the single-stage items, the book simply asked subjects to agree predictions and explanations.

Procedure

(a) Pre-test

For the pre-test, the children were taken individually into a vacant classroom. After a brief introduction, the interviewer set up the standard display, and presented the items orally. When presenting the three-stage items, the interviewer always altered the apparatus as required by the schedule before asking the questions. Once the children had answered, they were occasionally allowed to roll the vehicles down. This was purely in the interests of interviewer–subject harmony and, to avoid influencing conceptions prior to the group task, it was only permitted when the correct area had been predicted. The interviewer recorded the children's responses in note form during the pre-test, and at the end indicated an assessment of their English. Five children were excluded from further participation because their English was deemed inadequate.

(b) Scoring and grouping

The responses made by the remaining children were used to assess their alternative conceptions. Assessment began by identifying the explanations given first for the angle manipulations, then for the starting position, then for the surface friction and finally for the object weight, and scoring these

Table 1 Principles of scoring

Score	*Angle/starting position/surface friction*	*Object weight*
1	Variable not considered or confusion about how it operates, e.g. increasing angle or decreasing surface friction will decrease distance travelled.	Variable not considered.
2	Understanding of how variable operates but inability to coordinate variable with another, e.g. increasing angle or decreasing surface friction will increase distance travelled.	Variable believed to be important but not coordinated with another variable, e.g. increasing weight will increase distance travelled.
3	Understanding of how variable operates and coordination with another variable, e.g. increasing angle will reduce the effects of increasing surface friction.	Variable believed to be important and coordinated with another variable.
4	Full understanding of how variables coordinate, e.g. distance travelled is directly related to starting position height, increasing angle will decrease the effects of surface friction when the latter is held constant.	Variable excluded as irrelevant, e.g. object weight makes no difference.

with reference to Table 1. To check reliability, the responses from a randomly chosen 25 per cent of the pre-tests were scored by two judges. Their agreement was 87.9 per cent.

Using the scores, 27 children were categorized as 'Level I'. These children not only failed to coordinate, scoring 1 or 2 for at least 50 per cent of their responses. They were also uncertain about how one or more of the relevant variables operated, scoring 1 for at least 50 per cent of their responses to angle, starting position and/or surface friction. A further 44 children were categorized as 'Level II'. These children also failed to coordinate when judged by the above criterion. However, they were clear about how the relevant variables operated, scoring 2 or more for at least 50 per cent of their responses to angle, starting position and/or surface friction. A total of 37 children were categorized as 'Level III'. These children scored 3 or more (though a score of 4 was rare) for at least 50 per cent of their responses to all four variables, indicating clarity about how the factors operate and some coordination. Consistent with the results of Ferretti *et al.* (1985) and Inhelder & Piaget (1958), there was some tendency for level to increase with age band, although this was not statistically significant ($\chi^2(6) = 11.68$, n.s.).

Table 2 Groups as a function of pre-test level

		Differing		*Similar*
Low	Low D	Six groups each containing two Level I children and two Level II children (i.e. $N = 24$)	Low S	Three groups each containing four Level I children (i.e. $N = 12$) Three groups each containing four Level II children (i.e. $N = 12$)
High	High D	Six groups each containing two Level II children and two Level III children (i.e. $N = 24$)	High S	Three groups each containing four Level III children (i.e. $N = 12$)

Using the ascribed levels, the children were grouped into foursomes as shown in Table 2. In forming the groups, steps were taken to ensure, firstly, that the members of each group came from the same school class and, secondly, that the members of each D group differed as much as possible in ways apart from level while the members of each S group differed as little as possible. Thus, the D groups always had two children who thought that heavy vehicles would travel further, and two who thought that light vehicles would do this. The S groups were always homogeneous over object weight. The low D groups always had Level I children who differed over the variable/s of which they were uncertain. The low S groups always had Level I children who were similar.

Such considerations meant that some pre-tested children had to be excluded from the group task. Had age and sex also been considered, subject wastage would have become acute. Accordingly, these factors were ignored. Despite this, there were no significant sex differences between the low D children and the low S ($\chi^2(1) = 2.12$, n.s.) nor between the high D children and the high S ($\chi^2(1) = 0.08$, n.s.). Equally, there was no significant age difference between the low D children and the low S ($t(46) = 0.83$, n.s.). The age difference between the high D children and the high S was, however, statistically significant ($t(46) = 2.56$, $p < .05$), with the high D children being on average 7 months younger than the high S.

(c) Group task

The group task was presented by an experimenter who had not been involved in either the pre-testing or the scoring and grouping, and who was ignorant

of the type of group at the time of the task. This experimenter took the children in their groups to the classroom used for pre-testing, reassured them about a video-camera that was recording throughout, and explained the task. She then set up the standard display. Once the car had come to rest, she gave each child a set of cards, and invited them to make the predictions. The need to work independently was emphasized, and the experimenter demonstrated the display referred to on the cards (without, of course, rolling the vehicles) prior to each prediction.

Once the predictions had been made, the experimenter produced the book, and took the children through the text until they had completed the second three-stage item. She did not give feedback on the decisions, but checked that the reading was manageable and the procedure (especially the need to discuss and agree) had been grasped. For subsequent items, the children were on their own. When they had finished, the experimenter returned and, once more without giving feedback, enquired about some of the decisions. In total, the group task lasted between 45 and 75 minutes.

(d) Post-test

Prior to the task, one child in each group had been randomly chosen for the immediate post-test. This child was given a set of cards marked with a sticker. At the end of the task, the children were asked to look for the sticker, and the 'winner' invited to 'do the task again'. Since most children were disappointed to lose, it was clear that enthusiasm for the task was in no sense diminished by its lengthy duration.

The immediate post-test was presented the afternoon following a morning group task or the morning following an afternoon one. It kept to the same procedure as the pre-test, except that it was conducted by the group task experimenter. The pre-test interviewer did, however, present the delayed post-test which was administered to all 84 group participants. The procedure for the delayed post-test was the same as for the pre-test and the immediate post-test. Scoring of the immediate and delayed post-tests was done in ignorance of the children's groups, and, like the pre-test, was with reference to Table 1.

Results

It will be remembered that one aim of the study was to see whether grouping such that alternative conceptions differ has the beneficial effects with motion down an incline that Howe *et al.* (1990) reported for flotation. These beneficial effects were in terms of the progress that individual children made towards the received wisdom of science when they were tested some weeks after a group task. Thus, it was learning in an individual and not necessarily immediate sense that was the primary concern, and in the context of the

present study, this meant focusing on learning as defined by pre- to delayed post-test change. Before proceeding, however, it was necessary to decide whether the pre- and delayed post-test scores would have to be analysed separately for each variable or whether they could be combined across variables. Accordingly, the mean scores for angle, then starting position, then surface friction and finally object weight were computed for each child's pre- and delayed post-test. Each pre-test mean was subtracted from the corresponding delayed post-test mean to produce a measure of change. Seven one-way ANOVAs were carried out to see whether the amount of change differed between variables for, respectively, the Level I children in the low D groups, the Level I in the low S, the Level II in the low D, the Level II in the low S/high S, the Level II in the high D, the Level III in the high D and the Level III in the high S. The results for the Level II children in the low S/high S groups proved significant ($F(3,33) = 5.08$, $p < .01$) with more change for angle and surface friction than for starting position or object weight. However, as there were no other significant results, a composite measure seemed warranted. Therefore, the means across, firstly, all pre-test scores and, secondly, all delayed post-test scores were computed for each child, and the former subtracted from the latter as the measure of change.

Once computed, the scores were organized as indicated by Table 3, and level × condition ANOVAs carried out on first the low children and then the high. By separating the analyses in this fashion, it was possible to avoid problems resulting from having the same Level II children in both the low S and the high S groups. With the low children, there was no significant level effect and no significant interaction, but there was a significant condition effect ($F(1,44) = 10.25$, $p < .01$). Thus, regardless of whether they had started at Level I or Level II, the children in the D groups progressed more. This was of course consistent with benefits accruing from grouping such that alternative conceptions differ. With the high children on the other hand, there was a significant level effect ($F(1,44) = 14.67$, $p < .001$), but there was neither a significant interaction nor a significant condition effect. Here then, regardless of condition, the Level II children progressed more. The absence of a significant condition effect cannot have been an artefact of the age difference between the conditions documented earlier. The correlations between pre- to delayed post-test change and age were +.05 (n.s.) and −.14 (n.s.) for the high D and the high S children respectively.

In addition to comparing the learning in the differing and similar groups, the study also had the aim of clarifying the process by which learning is effected. Its particular concern was whether learning could have been through the internalization of superior conceptions which the groups constructed jointly, and if not, when and how change was effected. To resolve the first issue, it was decided to analyse the group task interactions at the one point where the children were explicitly invited to construct joint conceptions, namely the point at which they were asked to agree explanations of outcomes

Table 3 Change from pre-test to delayed post-test

	Mean pre-test score[a]	*Mean delayed score*[a]	*Mean change pre- to delayed post-test*
Low D children			
Level I	1.91 (.25)	2.52 (.38)	+.61
Level II	2.34 (.18)	2.84 (.11)	+.50
All low D	2.12 (.30)	2.68 (.32)	+.56
Low S children			
Level I	2.02 (.28)	2.38 (.28)	+.36
Level II[b]	2.40 (.13)	2.69 (.19)	+.29
All low S	2.21 (.29)	2.54 (.29)	+.33
High D children			
Level II	2.40 (.08)	2.72 (.21)	+.32
Level III	2.71 (.16)	2.78 (.18)	+.07
All high D	2.56 (.20)	2.75 (.20)	+.19
High S children			
Level II[b]	2.40 (.08)	2.69 (.21)	+.29
Level III	2.67 (.13)	2.82 (.17)	+.15
All high S	2.54 (.19)	2.76 (.19)	+.22

[a] SD in parentheses.
[b] Low S Level II children ≡ high S Level II children.

that were at variance with their predictions. It was recognized that the children were not precluded from constructing joint conceptions at other points. However, preliminary scrutiny of the videotapes had revealed that when joint constructions occurred, it was only at the explicitly signposted points.

The issue under scrutiny seemed to imply two separate questions: (1) how superior are the explanations that individual group members construct? and (2) how many other group members accept each explanation? Accordingly, the relevant interactions were located on the videotapes, and an attempt was made to identify the explanations to which each child was subscribing. These explanations were then scored using Table 1 and a count was made of the number of other group members by whom they were accepted. It was not always easy. The children did not advance explanations at every stage, and (despite the utilization of verbal and non-verbal information) it was not always clear who was accepting and who was not. For purposes of analysis, ambiguous cases were discarded, and when what can be called mean 'within-group performance' and 'number of agreements' scores were computed for each group member, it was the remaining instances that were considered. The way this operated in practice can be clarified with reference

Table 4 Correlates of within-group change

	Mean within-group change	*Mean number of agreements*	*Correlations between within-group change and number of agreements*	*Correlations between within-group change and pre- to delayed post-test change*
Low D groups				
Level I	+.14	1.62	+.71**	+.45
Level II	−.18	1.44	+.52	−.01
All low D	−.02	1.53	+.59**	+.38
Low S groups				
Level I	+.11	1.96	−.11	+.56
Level II[a]	−.21	2.27	−.13	−.56
All low S	−.05	2.12	−.28	+.33
High D groups				
Level II	−.31	1.81	+.09	+.01
Level III	−.52	1.78	+.19	+.20
All high D	−.42	1.80	+.14	+.34
High S groups				
Level II[a]	−.21	2.27	−.13	−.56
Level III	−.51	2.35	−.77**	+.24
All high S	−.36	2.31	−.55**	+.22

** $p < .01$.

[a] Low S Level II children ≡ high S Level II children.

to Appendix II which presents interactions that contrast over both the number of agreements and the explicitness of the explanations. The scores were obtained by a single judge. However, to check her reliability, the children in four randomly chosen groups were independently scored by a second judge. The consensus between the two judges was 80 per cent over within-group performance and 66 per cent over number of agreements.

Treating the consensus as acceptable, the pre-test scores obtained by the group participants were subtracted from the within-group performance scores to produce measures of 'within-group change'. As Table 4 shows, the values were largely negative, indicating that, far from being superior to the initial conceptions, the conceptions elaborated in interpretation of outcomes were characteristically inferior. However, the children's within-group change scores were based on some conceptions that were accepted by other group participants and some that were not. Since, as Table 4 intimates, the overall level of acceptance was not particularly high, it is possible that the scores when conceptions were agreed were better than the scores when conceptions were not agreed. If this were the case, it might still be legitimate to argue for

Table 5 Immediate post-test related to pre-test and delayed post-test

Group	*Pre-test*	*Immediate post-test*	*Delayed post-test*
Low	2.20_a (2.17)	2.38_a	2.66_b (2.61)
High	2.55_a (2.54)	2.46_a	2.72_b (2.75)

Notes: Means in the same row whose subscripts differ are significantly different ($p < .05$). Unbracketed means are derived from the children who were given the immediate post-test. Bracketed means are derived from the whole sample.

jointly constructed conceptions being superior. To investigate further, within-group change was correlated with number of agreements. As Table 4 shows, the overall results were not encouraging. With the S children, the correlations were negative, suggesting that these children performed better when they failed to agree. With the D children, the correlations were positive, but they only reached statistical significance with the low D. Moreover, even here, there is little suggestion that the internalization of jointly constructed conceptions was involved in learning. As Table 4 shows, the correlations between within-group and pre- to delayed post-test change were never more than weakly positive. Given the generally regressive nature of within-group change coupled with the generally positive nature of pre- to delayed post-test change, this means that there must have been many children who advanced from pre- to delayed post-test despite group performances that were worse than their pre-test.

In view of these results, it would be hard to argue that learning involved the internalization of conceptions that were jointly constructed within the groups. However, this leaves unclear whether learning involved conceptions that were privately constructed at that time. In order to investigate this, mean scores across immediate post-test responses were obtained for the children who participated in this part of the study. The pre-test means were subtracted from these scores to produce measures of pre- to immediate post-test change. Correlations were calculated between pre- to immediate post-test change and pre- to delayed post-test change, firstly for the low children and secondly for the high. The small numbers of children receiving immediate post-tests precluded subdivision within the low and high groups. The correlations were +.53 ($p < .1$) for the low children and +.63 ($p < .05$) for the high, suggesting that private construction while the groups were in progress may have been relevant. This accepted, it was unlikely to be the whole story as can be seen from the mean pre-test, immediate post-test and delayed post-test scores shown in Table 5. These scores were compared using one-way ANOVAs. As they proved significant ($F(3,33) = 12.36$, $p < .001$ for the low children, and $F(3,33) = 18.83$, $p < .001$ for the high), *post hoc*

comparisons were made using the Scheffé test (Kirk, 1968). As Table 5 makes clear, the immediate post-test scores did not differ significantly from the pre-test, but were significantly lower than the delayed post-test. This suggests that much of the progress took place once the group tasks were over.

Granted post-group progress, the issue is whether the crucial information was generated within the groups or solicited afterwards. The interviews at the end of the delayed post-tests suggest that it must have largely been the former. Only 19 children reported looking for further information, and for some the outcome was of dubious value. For example, 11 of the 19 claimed to have made direct observations, mostly via skateboarding or constructing slopes but one irrelevantly by varying the weights attached to balloons. Eight claimed to have consulted other people (mainly parents) but in one case this was to be told that object weight was critical. Four claimed to have read relevant books, but for one this was an account of car manufacture and for another it was the antics of Rudolph the Diesel! It is not then surprising that the children who reported looking for further information were no more likely than the other children to show above average pre- to delayed post-test change ($\chi(1) = .82$, n.s.).

Discussion

The starting point for the study was the alternative conceptions which, according to the literature, children aged 8 to 12 display over motion down an incline. With this age group, conceptions almost always include all the relevant variables and only one irrelevant one. Where there are differences is over how the relevant variables operate and whether these variables are coordinated into an integrated model. This provided the starting point for the study in that it contrasts with the alternative conceptions which children in the same age group display for flotation. Thus, it raised the question of whether research with motion down an incline would substantiate the evidence which Howe *et al.* (1990) provide for flotation that when children work in groups to discuss and test their alternative conceptions, progress is maximized when the conceptions differ. Finding out was one of the study's major aims and in the event, its results were mixed. They were consistent with Howe *et al.* in that the low D children showed significantly greater pre- to delayed post-test progress than the low S. They were inconsistent in that the high D children did not show significantly greater pre- to delayed post-test progress than the high S.

Seeking to explain the mixed results, two possibilities warrant attention. The first is that although working in groups where alternative conceptions differ does not invariably maximize learning, it is one of the conditions that must be fulfilled. The second is that although working in groups where alternative conceptions differ may be helpful, it is by no means necessary. Evidence for the first can be drawn from the fact, made clear by Table 3,

that besides failing to differ from the high S subjects, the high D children also progressed less than the low D. Of course, it could be argued that the high D performance was subject to ceiling effects or the tendency of relatively high scores to regress statistically towards the mean. However, this can only be part of the story. The mean pre-test score of the most advanced children in the high D groups, the Level III, was 2.71. Seeing from Table 3 that the mean pre- to delayed post-test change in the low D groups was only +.56, 2.71 seems sufficiently far from the ceiling of 4. In addition, the poor performance of the high D children relative to the low D was only partly caused by the Level III subjects. It was due also to the fact, apparent from Table 3, that the Level II children in the high D groups progressed less than the Level II children in the low D, a difference that was statistically significant ($t(22) = 2.50$, $p < .05$). This finding is of general theoretical interest because, being indicative of children learning more with lower performing peers than they did with higher, it is problematic for theories that rely on modelling. In the present context, it is strong evidence that the deflation of the high D performance resulted, in part at least, from a condition (or conditions) additional to differing conceptions which only the low D groups managed to meet.

The most obvious candidate for the condition/s is that the combination of conceptions reflects the low D groups rather than the high D. Perhaps, remembering what Level II plus Level I amount to, groups where alternative conceptions differ help when the combination involves differences over how the relevant variables operate. Perhaps, remembering what Level II plus Level III amount to, they do not help when the combination involves differences over whether the relevant variables are coordinated. If this were the case, extensive limits would be placed on the benefits to be gained from group composition. Assuming that, as children get older, the differences between them are increasingly likely to be over coordination, extrapolation from primary to higher educational levels would be rendered unsafe. Indeed, limits would also be indicated for the primary level itself, since research reported by Clough & Driver (1986), Kaiser, McCloskey & Profitt (1986), Piaget (1974) and Strauss (1981) suggests that the differences in 8- to 12-year-olds' conceptions of air, heat and free-fall are also partly in degrees of coordination. However, before the limits are taken as read, recent work by Thorley & Treagust (1987) needs to be noted. Working with science students at an Institute of Technology, these authors investigated the effects of group interaction on the understanding of mechanics and electricity. From the descriptions they provide, the interactions were almost certainly between individuals who differed over whether the relevant variables were coordinated, and they appear to have been beneficial.

Thorley & Treagust's data are far from conclusive, and in any event they relate to an older age group. Nevertheless, they do raise the possibility that

the additional condition/s may be something other than a combination of conceptions. Thinking what else might be involved, the requirement that the task have a particular form cannot be overruled. To see why, it should be noted that the differences within the low D groups were such as to guarantee disagreement over the predictions expressed on cards. After all, the predictions made by children who think steep angles inhibit are bound to conflict with those made by children who think steep angles help. However, as the predictions related to displays that changed one variable from the standard, the differences within the high D groups were not such as to guarantee disagreement over the predictions on cards. This meant that the high D children were not obliged by the group task instructions to discuss the predictions whereas the low D children were, and this may have been important. It is a conversational convention (Levinson, 1983) that when disagreement occurs, stances have to be justified and it is hard to think how the predictions in the present study could have been justified without reference to conceptions. Thus, in being obliged to discuss their predictions in a context of disagreement, the low D children were being offered an additional opportunity to interact over conceptions. The high D children were restricted to interaction in the context of jointly explaining outcomes. The difference may have been crucial, particularly when the examples of interaction over explanations already presented in Appendix II are compared with the lively exchanges over prediction differences in the following example:

(*The children had just put the heavy-weight vehicle at the middle starting position on the middle-friction slope inclined at the middle angle.*)

Moien [*reads from text*]: If you all ticked the same box, go on to the next page. If you did not, try to agree where the lorry will roll to.
Barnaby: I think it's the same square.
Imran: I think the same square, the same square.
Emily: But Moien did the further square. I did the same square.
Barnaby: I think it's because it's not on a steep slope.
Moien: But it's heavy on it.
Barnaby: I think it'll go to the middle one.
Emily: Shall we try it?
Barnaby: No we've got to agree. I think it'll go to the middle one. It'll go to the middle one because it's not on a very steep slope.
Moien: But it's got weight on it.
Imran: It's the slope that's important, and where it starts from.
Barnaby: The weight doesn't matter.
Emily [*to Moien*]: Will you change your mind?
Moien: I suppose so.

If discussion in the context of prediction disagreement was important, it would suggest a condition additional to group composition whose implications are not unduly restrictive. Tasks which guarantee contrasting predictions when the differences are over coordination could, after all, be readily designed. In addition, however, it would indicate that the resolution of conception disagreements was in no way essential. Earlier, it was pointed out that the 'joint construction' of conceptions only occurred after predictions had been tested, meaning that the resolution of conception differences cannot have taken place at the prediction stage. The example shows why this was. The children advanced conceptions of underlying variables to support their predictions, but they did not see the reconciliation of these conceptions as required for the agreement of predictions. However, once it is recognized that the centrality of prediction formulation implies the non-centrality of conception resolution, the question is raised as to whether further evidence can be found for the latter. It probably can be when it is noted from Table 4 that the 'number of agreements' scores were not very high. Remembering that these scores were computed from interactions after predictions were tested, it can be inferred that even though joint construction did occur at this stage, it was still infrequent, meaning that even here there was some considerable failure to resolve conception disagreements. Yet failure to agree cannot have inhibited learning, particularly when, as Table 4 shows, the mean number of agreements was lowest with the low D children who learned the most. Indeed, the low D children produced significant positive correlations between number of agreements and within-group change but insignificant correlations between within-group change and pre- to delayed post-test change. This also suggests that it was not the resolution of conception disagreements that mattered for growth.

Of course, if the resolution of conception differences is by no means essential, it follows that learning cannot have proceeded by the internalization of conceptions which the groups jointly constructed. However, this notion was found wanting on other scores. It was not just the low D children who produced insignificant correlations between within-group and pre- to delayed post-test change. It was all the children. Moreover, with the low D children as indeed with the others, overall within-group change was regressive while pre- to delayed post-test change was positive. Thus, there are additional reasons for concluding that learning cannot have involved the internalization of jointly constructed conceptions, and this is of course important. It is, as intimated earlier, contrary to what Doise & Mugny (1984) appear to imply. Moreover, it is also problematic for theorists who look to Vygotsky (e.g. 1987) for an all-embracing analysis of development and learning, for here too the notion of internalization plays a crucial role. It is true that Tudge (1990) departs from Vygotsky in proposing that what is internalized from social interaction will not necessarily be progressive. It is also true

that Forman (1989) sees the internalization process as mediated and perhaps undermined by decontextualization. Nevertheless, these writers share with Vygotsky the conviction that when a child operates independently 'he continues to act in collaboration . . . This help—this aspect of collaboration—is invisibly present. It is contained in what looks from the outside like the child's independent solution of the problem' (Vygotsky, 1987, p. 216).

Of course, rejecting internalization in the context of the present study does not entail rejecting it in every context. Garton (1984) has pointed out that most work in the Vygotskian tradition is concerned with practical problem solving. It may be, as Forman & Cazden (1985) intimate, that internalization operates in practical contexts but not in conceptual. This is certainly an issue for further research. Pending such research, it should be noted that, working with conceptual topics in non-science domains, Emler & Valiant (1982), Mackie (1980) and Roy & Howe (1990) have obtained results that are also hard to reconcile with an internalization process. Thus, there are clearly some areas where the effects of group interaction cannot be by way of internalization, and it is appropriate to look for an alternative process. From the results of the present study, the most plausible candidate seems to be a process which involves the private resolution of conflicts between conceptions made salient by the group interactions. The fact that pre- to immediate post-test change was positively correlated with pre- to delayed post-test change, coupled with the fact that learning seems to have been largely on the basis of within-group information, seems to signal the impact of group experiences. At the same time, the fact that pre- to immediate post-test change was less than pre- to delayed post-test change suggests that the experiences created conflicts to be resolved rather than solutions to be remembered. What the indications are, then, is a learning process which makes conceptual growth implicit, and this of course also concurs with the notion that discussion in the service of prediction resolution is all important. However, regarded more generally, the implied process squares equally with the emphasis which, as noted earlier, Piaget (1985) places on 'internally experienced conflicts' and gradual equilibration, suggesting that in some contexts at least such notions remain relevant to developmental theory. Finally, to conclude with the educational issue with which the paper began, the process would, if similar effects were found in ordinary classroom contexts, have profound implications for the pacing of teaching.

Acknowledgements

The research reported in this paper was supported by ESRC grant C00232426. Thanks are due to the ESRC and also to the schools who participated in the research and its pilot study.

References

Ames, G. J. & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology*, **18**, 894–897.

Berkowitz, M. W., Gibbs, J. C. & Broughton, J. M. (1980). The relation of moral development to developmental effects of peer dialogues. *Merrill-Palmer Quarterly*, **26**, 341–357.

Champagne, A. B., Gunstone, R. & Klopfer, L. E. (1983). Effecting changes in cognitive structure amongst physics students. Paper presented to American Educational Research Association, Montreal.

Clough, E. E. & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education*, **70**, 473–496.

Damon, W. & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly*, **28**, 347–367.

Department of Education & Science (1989). *Science in the National Curriculum*. London: HMSO.

Doise, W. & Mugny, G. (1984). *The Social Development of the Intellect*. Oxford: Pergamon.

Driver, R., Guesne, E. & Tiberghien, A. (1985). *Children's Ideas in Science*. Milton Keynes: Open University Press.

Emler, N. & Valiant, G. (1982). Social interaction and cognitive conflict in the development of spatial co-ordination skills. *British Journal of Psychology*, **73**, 295–303.

Ferretti, R. P., Butterfield, E. C., Cahn, A. & Kerkman, D. (1985). The classification of children's knowledge: Development of the balance-scale and inclined-plane tasks. *Journal of Experimental Child Psychology*, **39**, 131–160.

Forman, E. A. (1989). The role of peer interaction in the social construction of mathematical knowledge. In N. Webb (Ed.), *Peer Interaction, Problem Solving and Cognition: Multidisciplinary Perspectives*. Oxford: Pergamon.

Forman, E. A. & Cazden, C. B. (1985). Exploring Vygotskian perspectives in education: The cognitive value of peer interaction. In J. V. Werstch (Ed.), *Culture, Communication and Cognition: Vygotskian Perspectives*. Cambridge: Cambridge University Press.

Forman, E. A. & Kraker, M. J. (1985). The social origin of logic: The contribution of Piaget and Vygotsky. In M. W. Berkowitz (Ed.), *Peer Conflict and Psychological Growth*. San Francisco: Jossey-Bass.

Garton, A. F. (1984). Social interaction and cognitive growth: Possible causal mechanisms. *British Journal of Developmental Psychology*, **2**, 269–274.

Hewson, M. G. & Hewson, P. W. (1983). The effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching*, **20**, 731–743.

Howe, C. J., Tolmie, A. & Rodgers, C. (1990). Physics in the primary school: Peer interaction and the understanding of floating and sinking. *European Journal of Psychology of Education*, **V**, 459–475.

Inhelder, B. & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.

Kaiser, M. K., McCloskey, M. & Profitt, D. R. (1986). Development of intuitive theories of motion. *Developmental Psychology*, **22**, 67–71.

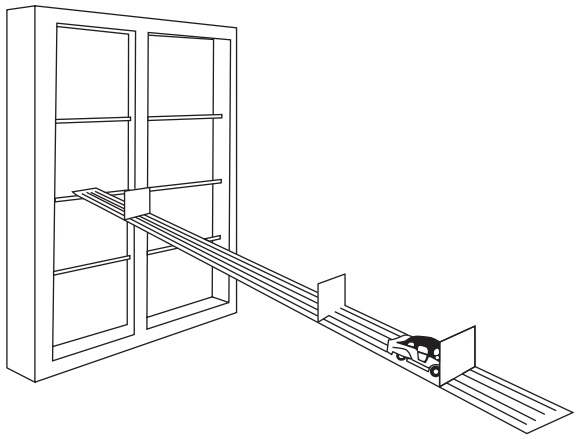
Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
Kirk, R. E. (1968). *Experimental Design Procedures for the Behavioural Sciences*. Belmont, CA: Brooks Cole.
Mackie, D. (1980). A cross-cultural study of intra-individual and inter-individual conflicts of centrations. *European Journal of Social Psychology*, **10**, 313–318.
McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today*, **37**, 24–32.
Nussbaum, J. & Novick, S. (1981). Brainstorming in the classroom to invent a model: A case study. *School Science Review*, **62**, 771–778.
Osborne, R. & Freyberg, P. (1985). *Learning in Science*. Auckland: Heinemann.
Piaget, J. (1930). *The Child's Conception of Physical Causality*. London: Routledge & Kegan Paul.
Piaget, J. (1932). *The Moral Judgment of the Child*. London: Routledge & Kegan Paul.
Piaget, J. (1972). *The Principles of Genetic Epistemology*. New York: Basic Books.
Piaget, J. (1974). *Understanding Causality*. New York: Norton.
Piaget, J. (1985). *The Equilibration of Cognitive Structures*. Chicago: Chicago University Press.
Roy, A. W. N. & Howe, C. J. (1990). Effects of cognitive conflict, socio-cognitive conflict and imitation on children's socio-legal thinking. *European Journal of Social Psychology*, **20**, 241–252.
Stead, K. & Osborne, R. (1981). What is friction?—Some children's ideas. *Australian Science Teachers' Journal*, **27**, 51–57.
Strauss, S. (1981). *U-shaped Behavioural Growth*. New York: Academic Press.
Thorley, N. R. & Treagust, D. F. (1987). Conflict within dyadic interactions as a stimulant for conceptual change in physics. *International Journal of Science Education*, **9**, 203–216.
Tudge, J. (1990). Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice. In L. C. Moll (Ed.), *Vygotsky and Education: Instructional Implications and Applications of Sociohistorical Psychology*. New York: Cambridge University Press.
Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds), *The Collected Works of L. S. Vygotsky*. New York: Plenum.
Weinstein, B. D. & Bearison, D. J. (1985). Social interaction, social observation and cognitive development in young children. *European Journal of Social Psychology*, **15**, 333–343.

Appendix I: Extract from group task text

(The extract gives the text for the *first* stage of the *first* item. The text became progressively briefer to avoid labouring instructions that are well understood.)

The lowest gate

To start off, close the lowest gate on slope B and put the car behind it. Now, each of you must find your Card 1. Do this before reading on.



Have you done this? If you have, look at what each of you has ticked. Did you all think that the car would roll to the *nearer* square? If so, go on to the next page. If not, read on.

Did all of you think that the car would roll to the *same* square as the other car? If so, go on to the next page. If not, read on.

Did all of you think that the car would roll to the *further* square? If so, go on to the next page. If not, read on.

Did some of you think that the car would roll to the *nearer* square, and some of you think that it would roll to the *same* square as the other car or the *further* square? Look at the car together, and talk about which square the car will roll to. When you have agreed, go on to the next page.

Have you agreed? When you are ready, pull the gate up so that the car can roll down the slope. Watch carefully to see where it stops.

What happened? Did things turn out the way you all thought? If so, go on to the next page. If not, read on.

Talk very carefully about what happened. Try to agree why the car stopped where it did. Make sure that everybody in the group says what they think. Then talk about the different ideas until you agree which are right. Take your time and do not go on until you all think the same way.

Do you all agree why the car stopped in that square? If you do, turn to the next page.

Appendix II: Group task scoring

High agreement

(*The children had been incorrect regarding the heavy-weight vehicle on the low-friction slope with the middle starting position and the middle angle*.)

George (*reads from text*): Try to agree why the lorry stopped where it did. Then go on to the next page.
David: We can't go on to the next page then. We've got to agree.
Sam: It's because it's got a smooth track.
David: 'Cos it's got a smooth track.
Others: Yes.
Sam: Good on you all, you agreed with me.

(Each child was awarded a 'within-group performance' score of 2 for understanding how surface friction operates without coordination and a 'number of agreements' score of 3 for seeming to concur with everyone.)

Low agreement

(*The children had been incorrect regarding the heavy-weight vehicle on the middle-friction slope with the high starting position and the low angle*.)

Andrew (*reads from text*): If the lorry stopped where you thought, go on to the next page. If not, try to agree why it stopped where it did. Do not go on until you all agree.
Abrar: It's because there's not much hill.
Sirinder: Because I got it correct.
Kemal: No you didn't because you agreed with us.
Sirinder: No I didn't. It's because you made me agree.
Kemal: No I didn't.
Others: Yes you did.
Andrew: It's because it's a long slope.
Abrar: It's because it's not much of a hill. The peg's down.
Sirinder: I agree.
Kemal: I thought it would go further.

(Andrew was awarded a 'within-group performance' score of 2 for recognizing the relevance of starting position without coordination. His 'number of agreements' score was 0 in that nobody seemed to concur with him. Abrar and Sirinder were also awarded 'within-group performance' scores of 2, this time for recognizing the relevance of angle without coordinating. Their 'number of agreements' scores were 1 for concurring with each other. Kemal was too inexplicit to be coded.)

75

ON THE COMPLEX RELATION BETWEEN COGNITIVE DEVELOPMENTAL RESEARCH AND CHILDREN'S SCIENCE CURRICULA

K. E. Metz

My earlier article (Metz, 1995) identified several assumptions about elementary school children's scientific reasoning abilities that have frequently been used for the purpose of framing "developmentally appropriate" science curricula. That article traced the origin of those assumptions to an interpretation of a segment of Piaget's writings and then critiqued those assumptions of the basis of Piaget's corpus, as well as the contemporary cognitive developmental research literature. Given that developmental research constituted the primary base on which I critiqued these assumptions and formulated alternative recommendations, I am surprised by Deanna Kuhn's (1997) contention that the article could be read as suggesting that the developmental literature has "failed" science educators and that they would be advised to look elsewhere to inform their curricular design. Nevertheless, I do consider the relation between cognitive developmental research, as embodied in the contemporary research tradition, and children's science curricula as fundamentally complex. This essay examines three interrelated characteristics of the cognitive developmental research tradition that contribute to the complexity of this relationship: (a) its tendency to attribute shortcomings in performance to the child's stage, with the assumption that these shortcomings will disappear with sufficient advancement of cognitive development: (b) the frequent confounding of weak knowledge with developmentally based cognitive deficiencies; and (c) the emphasis of robust stage-based constraints on children's thinking, to the neglect of variability and change.

Source: *Review of Educational Research*, 1997, 67(1), 151–163.

In an earlier article (Metz, 1995), I critiqued three assumptions about elementary school children's scientific reasoning abilities that have frequently be used in framing "developmentally appropriate" science curricula. In short, these assumptions were the following:

(1) Seriation and classification constitute the core intellectual strengths of elementary school children. Therefore, observation, ordering, categorization, and corresponding inferences and communications are appropriate science process objectives for children's science instruction.
(2) Elementary school children can comprehend only ideas that are linked to concrete objects, as they are "concrete thinkers." Therefore, educators should restrict children's science to hands-on activities and relegate abstract ideas to later grades.
(3) Not until adolescence do children grasp the logic of experimental control and inference. Therefore, educators should postpone scientific investigations, in the sense of design and implementation of experiments and drawing inferences from the complex of outcomes.

After tracing the genesis of these ideas to an interpretation of a small, albeit broadly circulated aspect of Piaget's work, the article analyzed their validity on the basis of both Piaget's large corpus and various contemporary cognitive developmental literatures—including the child's theory of mind developmental literature, the logical structures developmental literature, the scientific cognition developmental literature, and the cognitive science literature comparing scientific cognition of children and adults. It was on the basis of this review of research, the majority of which consisted of developmental works, that I concluded that these three assumptions are ill-founded and that this approach to children's science education significantly underestimates the potential of children's scientific reasoning abilities.

Above and beyond the failure of this instructional approach to capitalize on children's scientific reasoning abilities, my article identified other limitations in the design of science instruction along these lines. Most problematic, the targeting of purportedly elementary science processes for the first years of schools with a postponement of the integrated practice of goal-focused investigations until the higher grades results in decomposition and decontextualization in the teaching and learning of scientific inquiry. As a consequence, young children engage in science activities such as observation and categorization apart from a rich goal structure or overriding purpose, a practice which is detrimental from cognitive, motivational, and epistemological perspectives.

Although I argued that these three assumptions about the limitations of children's scientific reasoning are invalid, I made no claim that there do not exist significant limitations on children's reasoning in this sphere. My article relied on the developmental literature to identify the cognitive weaknesses

that are most fundamental and relevant to children's scientific inquiry: (a) metacognitive weaknesses, including difficulty in taking their knowledge or thinking as an object of thought, and (b) weaker domain-specific knowledge.

I concluded,

> Although the investigations of elementary school children will presumably be less sophisticated than those of adolescents or adults, due to children's more limited domain-specific knowledge and their weakness at thinking about their thinking, these differences do not negate the possibility of their posing questions, gathering and interpreting data, and revising their theories. The research literature supports the feasibility of a much richer framework for young children's science instruction, wherein the processes previously approached in the elementary school grades as ends in themselves become tools in a more contextualized and authentic scientific inquiry. (Metz, 1995, p. 121)

Given the fact that I based most of my argument on a broad range of developmental literatures, I am perplexed by Deanna Kuhn's (1997) assertion that "the message that one might take away from Metz's article is that science educators have tried developmental psychology and it has failed them" (p. 147). I argued that many science educators have greatly oversimplified and misinterpreted Piagetian theory, not that they erred in considering developmental theory in the design of science curricula.

Nevertheless, I do view the relation between developmental research and science education as complex, and it is the complexities of this relationship and needed elaborations of the cognitive developmental research agenda that I examine in the rest of this response. To anticipate my conclusions, I argue for the importance of more cognitive developmental research that differentiates relatively robust and immutable stage characteristics from malleable cognitive characteristics, in conjunction with an analysis of the experiences which affect these malleable characteristics; an extension that takes up the challenge of the muddy waters of development vis-à-vis learning.

Complexities of the relation between developmental research and children's science education

In their efforts to bring curricula into agreement with children's ways of knowing at different age levels, educators have frequently turned to the cognitive developmental literature. I found developmental literature a rich source—indeed, the most appropriate source—to challenge many science educators' assumptions about stage-based limitations on children's scientific reasoning, in that it provided strong evidence against the developmental assumptions frequently used by science educators in curricular design.

However, the relation between developmental research and science curriculum is far from straightforward.

One top-level issue concerns the appropriate use of cognitive developmental research in the conceptualization of curricula. William James, the 19th-century psychologist, described the relation in terms of establishing constraints. In his book *Talks to Teachers on Psychology*, James (1958) contended,

> You make a great mistake if you think that psychology, being the science of the mind's laws, is something from which you can deduce definite programmes and schemes and methods of instruction for immediate classroom use. . . . A science only lays down lines within which the rules of the art must not transgress. Everywhere the teaching must agree with the psychology, but need not necessarily be the only kind of teaching that would so agree. (pp. 23–24)

Indeed, many contemporary educators use developmental theory to derive "lines" or constraints for framing age-appropriate curricula, as reflected in the set of constraints underlying children's science instruction that I identified and critiqued in my earlier article.

Nevertheless, there are a number of characteristics of the cognitive developmental research tradition that complicate the derivation of lines for framing age-appropriate curricula and that make problematic the relatively straightforward principle of agreement, articulated by James, between developmental findings and educational programs. This essay examines three related characteristics of the cognitive developmental research tradition and the complications they introduce in the use of cognitive developmental research to frame elementary science education: (a) the tendency to attribute shortcomings in performance to the child's stage, with the assumption that these shortcomings will disappear with sufficient advancement of cognitive development; (b) the frequent confounding of weak knowledge with developmentally based cognitive deficiencies; and (c) the emphasis of robust stage-based constraints on children's thinking, to the neglect of variability and change.

Attributing shortcomings to developmental stage

There exists a tendency in the cognitive developmental literature to attribute shortcomings in children's thinking to their developmental stage, with the assumption that the deficiency will resolve itself at a more advanced stage. This tendency is even stronger in educational translations of developmental theory. The sections below examine two examples of deficiencies that are frequently attributed to developmental shortcomings and yet also appear in the thinking of adults: (a) thinking tied to concrete and superficial features and (b) the failure to adequately differentiate theory from evidence.

As I argued in my earlier article, elementary school children's scientific reasoning is frequently characterized as concrete, in the sense that their knowledge of physical and biological phenomena focuses on the concrete features of objects, organisms, and phenomena. However, a review of the adult expert-novice literature reveals that concrete thinking of this form consistently characterizes the adult novice as well. Before constructing abstract knowledge of a given sphere, the novice—child or adult—is restricted to surface features. Thus Chi, Feltovich, and Glaser (1981) found that while experts categorized physics problems in terms of abstract principles, adults with little physics knowledge categorized physics problems at the level of surface features. Although the child tends to be a "universal novice" (Carey, 1985), studies that examine spheres in which children have deeper knowledge frequently document the possibility of abstract thought (e.g., Brown, 1990; S. A. Gelman & Markman, 1986).

The differentiation of theory and evidence constitutes a more complex illustration of a challenge to children and adults. The difficulty that elementary school children experience differentiating theory and evidence is frequently attributed to their stage of development. Thus the American Association for the Advancement of Science (AAAS, 1993), in its *Benchmarks for Science Literacy*, notes, "Research studies suggest that there are some limits on what to expect at this level of student intellectual development [Grades 3–5]. . . . Such students confuse theory (explanation) with evidence for it" (pp. 10–11). However, even adult scientists struggle with this distinction, albeit at a different level of sophistication. Philosopher of science Thomas Kuhn (1977) writes, "We [Popper and himself] both emphasize . . . the intimate and inevitable entanglement of scientific observation with scientific theory" (p. 267). Similarly, Stephen Toulmin (1972) has argued,

> Our own interest in facts is always to discover what can be made of them in light of current ideas. . . . In the solution of conceptual problems, the semantic and the empirical elements are not so much wantonly confused as unavoidably fused. (p. 189)

The case of Darwin's construction of the theory of natural selection illustrates the complex relation between theory and evidence on three levels: Darwin's thinking, the thinking of ornithologists of Darwin's and more recent generations, and the thinking of Darwinian scholars who have sought to understand the genesis of his theory. According to the textbook account, Darwin's observations and interpretation of variability among finches during his voyage through the Galapagos Islands constituted the beginnings of his theory of natural selection. From this perspective, Darwin's mid-voyage interpretation of modifications in finch beak size, to the point of the development of new and different finch species on the different islands, constituted a key impetus for the development of his theory. However, recent scholarship

analyzing the role of Darwin's finches in the development of his theory indicates otherwise.

When Darwin was recording his observations on the Galapagos Islands, he did not systematically note the islands on which he found the birds in his collection, but in many cases simply labeled their location as "Galapagos Islands" (Gould, 1980; Sulloway, 1982). Furthermore, analysis of his notes from onboard ship indicate that during his travels through the Galapagos Islands, Darwin thought about the inter-island variability of mockingbirds and tortoises, not finches (Sulloway, 1982). (The vice governor of the islands told Darwin he could tell which island a tortoise came from by its form, and Darwin appears to have independently noted two forms of mockingbirds living on different islands.) Indeed, it was only after he returned to England, where London ornithologist John Gould identified many of his specimens as different species of finches, including new species, that Darwin realized that the finches and the specific islands on which they were found constituted relevant data for his emergent theory of natural selection (Gould, 1980; Sulloway, 1982).

Subsequent attempts to reconstruct and then "correct" the island locality data reveal this same entanglement. Sulloway (1982) has documented Darwin's attempts to construct the locality data through correspondence with the two other individuals on his ship who had collected finches in the Galapagos Islands: Darwin's personal servant and FitzRoy, the ship's captain. Unlike Darwin, FitzRoy had consistently recorded the island from which he collected each specimen. Ironically, following publication of Darwin's *The Zoology of the Voyage of the H.M.S. Beagle* (1841), ornithologists changed the island locale data for FitzRoy's specimens in the British Museum to agree with Darwin's theory that each of the new finch species occurred only on a single island—a theoretical premise that later proved to be erroneous (Sulloway, 1982). Thus the British Museum curators first mistrusted their data, because the data conflicted with the prevailing theory, and then changed the data, purportedly to make it more accurate.

The complicated relation between theory and evidence is also reflected in the extent to which the particular methodology applied restricts and constrains the data one collects, as well as one's conception of the phenomenology under study. During the Beagle's voyage (1831–1836), Darwin typically collected a few specimens of each species, an approach that Sulloway (1982) attributes to "the typological and creationist assumptions that he brought with him to that archipelago" (pp. 18–19). This method is poorly suited to developing a database for the study of variation and change. Such a theoretical focus requires large samples. From this perspective, consider Darwin's large sample observations, at a point when he was struggling with the genesis and mechanism of variation: "Saw in Loddiges garden 1279 varieties of roses!!! proof of capability of variation" (as cited in Gruber, 1981, p. 159). Gruber contends that this observation reflects Darwin's

recognition of the huge magnitude of the variability and foreshadows his fundamental theoretical shift from the idea of environmentally induced, teleological variations to random variations.

Detailed analyses of Darwin's notebooks indicate that his coming to see his data in terms of the theoretical perspective of natural selection was supported by his wide reading in other disciplines, including politics and economics (e.g., Gruber, 1981, Kohn, 1980; Schweber, 1977). Ironically, Darwin's post hoc reconstruction of his theory's genesis as a largely empirical endeavor of "without any theory, collect[ing] facts on a wholesale scale" (as cited in Gould, 1980, p. 61) smacks of Thomas Kuhn's (1962) parodying phrase, the dogma of "immaculate perception." Darwin's statement concerning the genesis of his theory may in turn be attributed to the Baconian lens of his zeitgeist.

The fact that many philosophers and historians of science emphasize the "entanglement" of theory and evidence undermines any simple attribution of children's shortcomings in this sphere to developmental stage. Nevertheless, the existence of a nondevelopmental component does not negate the possibility of a developmental aspect. Thus, in the case of this particular shortcoming, the challenge that children experience in thinking about thinking (Brown, 1987) would presumably complicate the nontrivial task of differentiating theory and evidence.

The complex interaction of developmental and nondevelopmental factors in age-correlated shortcomings complicates science educators' use of cognitive developmental theory in the derivation of children's science programs. If educators assume that a particular weakness in children's thinking will automatically disappear at later stages of development, the tendency will be to forgo consideration of how the weakness might be ameliorated. Thus, children's science instruction has been frequently designed to avoid abstract thought rather than to strengthen it, and the challenging relation of theory and evidence has typically been left for higher grade levels. A warning in the *Benchmarks for Science Literacy* (AAAS, 1993) seems particularly important in this regard. Concerning limitations typically attributed to developmental stage, such as weak experimental design or the confusion of theory and evidence, this work cautions, "The studies say more about what students at this level do not learn in today's schools than about what they might possibly learn if instruction were more effective" (p. 11).

In summary, science educators cannot assume that age characteristics are simply a function of development in the sense of immutable cognitive characteristics of the stage. While some age-correlated weaknesses may be fairly robust at a particular stage and readily ameliorated at a subsequent stage, other weaknesses may to varying degrees respond to instruction, and still others may constitute an enduring challenge at all ages and all levels of expertise. For the advancement of both cognitive developmental theory and instructional practice, we need a research base that more adequately makes these distinctions.

Confounding weak knowledge with developmentally based cognitive deficiencies

A huge complicating factor in the use of cognitive developmental theory to guide children's science instruction is the fact that this research tradition has frequently ignored the influence of domain-specific knowledge in the design of experimental procedures and analysis of results. Consequently, children's weak knowledge has repeatedly been confounded with inadequacy of cognitive processing.

The confounding of weak knowledge with developmentally based weak information processing has been particularly pronounced in the school-age cognitive developmental literature, a tendency that Ann Brown (1990) has attributed to the influence of Piaget's structuralist orientation. Susan Carey's (1985) examination of how Inhelder and Piaget's (1955/1958) influential book *The Growth of Logical Thinking From Childhood to Adolescence* confounded children's inadequate domain-specific knowledge (such as the differentiation of weight, size, and density) with weaknesses in children's apparent "logic of inquiry" provides a lovely illustration of this problem. Ironically, the preschool cognitive developmental literature, which has generally paid keen attention to issues of domain familiarity in the design of experimental procedures, frequently portrays stronger competence of its subjects than the elementary school cognitive literature (Brown, 1990; Bullock, 1985; S. A. Gelman & Markman, 1986; Goswami & Brown, 1989).

Research comparing the performance of child domain-specific experts with adult novices again points to the fundamental importance of domain-specific knowledge. For example, Chi (1978) compared the abilities of child chess experts and adult chess novices to reconstruct from memory both chess boards that would normally appear in play and chess boards that would not. Although children have long been assumed to have a developmentally based shorter memory span, children outperformed the adults on their reconstruction from memory when the chess boards were ones that might well appear in a chess game. Chi concluded,

> The amount of knowledge a person possesses about a specific content area can determine to a large extent how well he or she can perform in both memory and metamemory tasks. The implication is that the sources of some of the age differences we often observe in developmental studies must be attributable to knowledge about the stimuli rather than to capacity and strategic factors alone. (p. 94)

The knowledge factor interacting with information processing has traditionally been conceptualized as domain-specific knowledge of a substantive sphere such as physics or even a sphere of physics. However, there are other

conceptualizations and forms of domain-specific knowledge that may well come into play. In particular, Brewer and Samarapungavan (1991) identify knowledge of the culture and accumulated methodological traditions of science as a critical factor supporting the adult scientist's theory building and, conversely, in its absence, undermining the child's. In short, we need to think beyond knowledge of specific subject areas in our analysis of the impact of knowledge on the adequacy of children's scientific reasoning.

The picture of competence manifested in any developmental study involves a complex interaction between cognitive development and the experiences of the children. In his thoughtful critique of the cognitive developmental research tradition, William Kessen (1984) argues, "A good part of what we call cognitive development is dependent on the selection by caretakers of possible lines of development in children" (p. 427). Similarly, Brown, Campione, Metz, and Ash (in press) contend that the findings of developmental studies, including work in children's theories of biological and physical phenomena, involve some unknown cultural and instructional component. Brown et al. assert,

> True to the tradition of this [cognitive developmental] discipline, cross-sectional data are taken from children divorced from the culture in which they are developing, a culture which includes school. We know a great deal about what the average (usually upper middle class) child knows about what is alive or not alive at age five, eight, ten, etc. What is not known, however, is the influence of instruction on these developmental milestones. (p. 26)

While several literatures provide strong evidence of the impact of knowledge on reasoning processes, other research focuses on the identification of biologically based stage characteristics. Susan Carey and Rochel Gelman's (1991) seminal book *The Epigenesis of Mind: Essays on Biology and Cognition* examines a number of contenders. Gallistel, Brown, Carey, Gelman, and Keil (1991) explain,

> Emlen . . . demonstrated that indigo buntings learn the constellations and the center of rotation of the night sky while nestlings—and only while nestlings. . . . The learning of stellar configurations by migratory songbirds illustrates the assumptions of domain-specific learning mechanisms. . . . This learning is specific to a particular developmental stage, even though what is learned is fundamental to important adult behaviors. Similarly, in humans, the learning of the phonetics of one's language community and some aspects of its grammar proceeds much more readily at a young age, even though what is then learned is used throughout adult life. . . . Secondly, the learning involves the operation of specialized computational

> mechanisms, dedicated to constructing a particular representation for a particular use. (pp. 17–18)

Of particular interest to science educators, Rochel Gelman (1990) discusses, as an example of what she views as biologically based processing mechanisms, how preschoolers learn about causality and the distinction between animate and inanimate objects. Gelman asserts that "young children [of 3 and 4 years of age] may be well on their way to developing a theory of action because they benefit from skeletal principles of causality that inform the processing mechanisms that respond to inputs that are relevant to animate and inanimate causality" (p. 102). However, Gelman is also careful to identify the importance of experience—for example, in attributing how children categorize to domain-specific perceptual mechanisms and their associated causal principles, as well as what the child has learned about the "predictive validity of cues" through domain-general learning mechanisms (p. 103).

In summary, the competence in scientific reasoning that children display at different age levels involves a complex interaction, largely unknown, of knowledge and developmental factors. It appears problematic to rely on this research base to derive curricular lines or constraints, in that it remains unclear which weaknesses are relatively immutable and which could be addressed by effective instruction.

Emphasizing stage-based constraints on children's thinking

There exists a widespread assumption that children's thinking develops in stages roughly corresponding to different age spans and that these stages are driven by robust constraints on children's thinking. Within this perspective, the issue of change is largely restricted to the study of transition from one multiyear stage to the next.

This orientation appears to have stemmed from Piaget's theory of sequential states in the emergence of cognitive structures, initially presumed to function as a kind of bedrock of limitations and abilities in children's reasoning. Its influence persists well beyond the scope of directly Piagetian-inspired research and, in simplistic form, is particularly prevalent in overviews of cognitive developmental theory for consumption by nonspecialists. In the words of cognitive developmentalist Robert Siegler (1994), "Most [developmental] theories place static states at center stage and change processes either in the wings or offstage altogether. . . . Thus, 10-year-olds [are said] to be incapable and 15-year-olds capable of scientific reasoning" (p. 1). Siegler and Shipley (1995) have argued,

> Although these 1:1 equations between ages and ways of thinking are omnipresent in the literature, few would defend them as literally

> meaning that young children of a given age or developmental level *always* use one approach, older ones *always* use another approach, and so on. Instead, their widespread and enduring use seems due to their having several pragmatic advantages. They are interesting, sometimes dramatic, easy to describe, easy to remember, and straightforward to discuss in textbooks and lectures. . . . The obvious problem is that 1:1 equations are inaccurate. . . . A less obvious but equally pernicious consequence of this oversimplification is that it impedes understanding of change. . . . Portraying children's thinking as monolithic at each point in the developmental sequence has the effect of segregating change from the ebb and flow of everyday cognitive activity. (p. 32)

In accordance with the traditional emphasis on identifying stage characteristics, much of the cognitive developmental research has utilized cross-sectional methodologies to get at snapshots of children's competence at different stages. In line with this agenda, most studies have employed experimental procedures involving a single session per subject, without any instructional component. Analysis of within-subject variability and change has frequently been omitted or deemed of secondary importance. This methodological lens has compounded the static stage-bound view of children's thinking, a view that drove the elaboration of these methodological traditions.

In response to a growing concern in Piagetian (Piaget, 1976) and information processing cognitive developmental communities (Sternberg, 1984) to better understand the process of change, methodologies such as microgenetic analysis are being elaborated that bring these issues to the fore. Through the lens of studies encompassing change (e.g., Karmiloff-Smith & Inhelder, 1974; Metz, 1985, 1993; Schauble, 1990; Siegler & Shipley, 1995), children's reasoning appears no longer "monolithic," but variable and dynamic.

Most of the cognitive developmental literature upon which elementary science education has derived curricular constraints has come from the classic tradition of stage description research, with negligible attention to variability or change. The agenda of identifying static stage characteristics diverges from the educators' focus on advancing the child's level of competence. The educator is deeply concerned with the issue of change. In particular, the question of the level of thinking children of a given age range could attain with effective instruction is at least as important as their level without instruction.

Vygotsky thought of these two perspectives on children's emerging competence as distinct conceptualizations of development. In his classic essay on the relationship between development and learning, Vygotsky (1978) argued,

> A well known and empirically established fact is that learning should be matched in some manner with the child's developmental level. . . .

> We must determine at least two developmental levels. The first level can be called the *actual developmental level*, that is, the level of development of a child's mental functions that has been established as a result of certain already *completed* developmental cycles. . . . *The zone of proximal development . . . is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*. (pp. 85–86)

Whether or not we choose to follow Vygotsky in conceptualizing the child's potential level of competence given the support of teachers and peers as a form of level of development, clearly there exists a significant gap between competence with and without such support. Furthermore, the zone of proximal development is of particular importance for the teacher and curriculum developer.

A research literature restricted to a study of actual developmental levels, as reflected in studies focused on snapshots of first-attempt levels of competence, bears a complex relation to instructional practice. As studies of actual developmental level can tell us only about children's thinking prior to instruction, they will presumably be useful in analysis of where instruction needs to begin, but less informative concerning the derivation of feasible instructional goals. Cognitive developmental research focused on cognitive change of the genre Siegler (1994) describes, as well as combinations of classroom-based and laboratory research that examine the possibilities of change from the perspective of long-term effective instructional interventions (Brown, 1992), constitutes a crucial research base for the design of developmentally appropriate science programs.

Conclusions

This essay examines several interrelated characteristics of the cognitive developmental research tradition that complicate its use as a base to frame developmentally appropriate science education for children. First, this research base tends to attribute shortcomings in performance to the child's stage, with the assumption that these shortcomings will disappear with sufficient advancement of cognitive development. Second, it frequently confounds weak knowledge with developmentally based cognitive deficiencies. Finally, it tends to emphasize robust stage-based constraints on children's thinking, to the neglect of variability and change.

Deanna Kuhn (1997) has suggested that science educators can more profitably use the developmental literature as a source of "guideposts" rather than constraints. I suggest that many of the same difficulties identified in this essay would challenge this form of application as well. Furthermore, I wonder to

what extent the interim states of competence—which I take to constitute Kuhn's guideposts—will change as a function of different powerful instructional programs, with different emphases and different teaching strategies.

We have no basis to assume that the impact of more effective science instruction will simply take the form of a speeding up, across the board, of the process of emergent scientific reasoning or that the trajectories of emergent competence will remain unchanged. Ann Brown and her colleagues (Brown et al., in press) have referred to multiyear instructional programs that "spiral" at increasingly complex levels around the same rich sphere of study as constituting a "developmental corridor." Their instantiation of this idea focuses on the field of environmental science, in conjunction with an emphasis on the fostering of scientific discourse and, more generally, reflective and analytic thought:

> We have introduced the term *developmental corridor* to capture the notion that units of FCL [their instructional project, Fostering a Community of Learners] should be revisited at ever increasing levels of complexity. This allows us to ask whether, after 4 or 5 years in the program, sixth graders will be capable of performing at much more mature levels of reasoning, capable of acquiring and using domain-specific knowledge of considerably greater complexity than the sixth graders in the program for the first time. In a very fundamental sense, to the degree FCL is successful we should be mapping a moving target. . . . Of considerable theoretical interest to developmental psychologists and of practical interest to designers of science curricula, are answers to the question: what, if any, forms of knowledge and process are immutable in the face of carefully tailored instruction?

Other innovative science instructional programs for children have other foci. For example, while Lehrer and Schauble's current project focuses on the development of children's model-based reasoning in science and mathematics, my project emphasizes children's design and interpretation of empirical research.

At issue here is what seemingly fundamental stage characteristics are modified by engagement in an excellent science program sustained across the elementary school years. Under such conditions, what would be our view of the development of children's scientific reasoning? What age-correlated characteristics of children's scientific cognition will remain constant, and what characteristics will change across contrasting forms of such instruction? Advancement of this research agenda has the power to strengthen both cognitive developmental theory and instructional theory.

Concerning cognitive developmental theory, this research agenda could address the issue of the immutable and the changeable at different stages across childhood. Given a rigorous protocol of parallel microgenetic laboratory

studies, wherein key developments are studied systematically at a fine grain of analysis, this agenda also has the potential to shed some light on the extraordinarily difficult question of the change process.

Concerning instructional theory, this research agenda could more adequately connect the spheres of cognitive developmental theory and instructional practice. In his chapter examining perils of the "marriage" between cognitive theory and instructional theory, a marriage that he claims frequently ends in divorce, Sternberg (1986) contends,

> Perhaps the single greatest source of disappointment in the application of cognitive principles to educational practice is the absence of an instructional theory to mediate the link between cognitive theory, on the one hand, and educational practice, on the other. (p. 378)

Theoretical advancements that more adequately differentiate immutable from mutable stage characteristics together with a rigorous analysis of the conditions of their mutability would empower cognitive developmental theory to more adequately inform instructional practice.

Acknowledgments

This work was, in part, supported by Research Grant No. RED-9453077 from the National Science Foundation. Opinions expressed are those of the author and do not necessarily reflect those of the foundation.

References

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Brewer, W., & Samarapungavan, A. (1991). Children's theories versus scientific theories: Differences in reasoning or differences in knowledge? In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives* (pp. 209–232). Hillsdale, NJ: Erlbaum.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other mysterious mechanisms. In F. E. Weinhert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Erlbaum.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science*, 14, 107–133.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141–178.

Brown, A., Campione, J., Metz, K. E., & Ash, D. (in press). The development of science learning abilities in children. In A. Burgen & K. Härnquist (Eds.), *Growing up with science: Developing early understanding of science*. Göteborg, Sweden: Academia Europaea.

Bullock, M. (1985). Causal reasoning and developmental changes over the preschool years. *Human Development*, *28*, 169–191.

Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 485–518). Hillsdale, NJ: Erlbaum.

Carey, S., & Gelman, R. (Eds.). (1991). *The epigenesis of mind: Essays on biology and cognition*. Hillsdale, NJ: Erlbaum.

Chi, M. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops*? (pp. 73–96). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(5), 121–152.

Gallistel, C. R., Brown, A. E., Carey, S., Gelman, R., & Keil, F. C. (1991). Lessons from animal learning for the study of cognitive development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 3–36). Hillsdale, NJ: Erlbaum.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, *14*, 79–106.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Goswami, U., & Brown, A. L. (1989). Melting chocolate and melting snowmen: Analogical reasoning and causal relations, *Cognition*, *35*, 69–95.

Gould, S. J. (1980). *The panda's thumb: More reflections in natural history*. New York: Norton.

Gruber, H. E. (1981). *Darwin on man: A psychological study of scientific creativity* (2nd ed.). Chicago: University of Chicago Press.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence* (E. A. Lunzer & D. Papert, Trans.). New York: Basic Books. (Original work published 1955)

James, W. (1958). *Talks to teachers on psychology*. New York: W. W. Norton.

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, *3*, 195–212.

Kessen, W. (1984). Construction, deconstruction, and reconstruction of the child's mind. In C. Sophian (Ed.), *Origins of cognitive skills: The Eighteenth Annual Carnegie Symposium on Cognition* (pp. 419–429). Hillsdale, NJ: Erlbaum.

Kohn, D. (1980). Theories to work by: Rejected theories, reproduction and Darwin's path to natural selection. *Studies in the History of Biology*, *4*, 67–170.

Kuhn, D. (1997). Constraints or guideposts? Developmental psychology and science education. *Review of Educational Research*, *67*, 141–150.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kuhn, T. S. (1977). *The essential tension*. Chicago: University of Chicago Press.

Metz, K. E. (1985). The development of children's problem solving in a gears task: A problem space perspective. *Cognitive Science*, *9*, 431–472.

Metz, K. E. (1993). Preschoolers' developing knowledge of the pan balance: From new representation to transformed problem solving. *Cognition and Instruction*, *11*(1), 31–93.

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, *65*, 93–127.

Piaget, J. (1976, June). Communication to the Symposium of the International Center for Genetic Epistemology, Geneva, Switzerland.

Schauble, L. (1990). Belief revision in children: The role of prior experience and strategies for generating evidence. *Journal of Experimental Child Psychology: Human Perception and Performance*, *11*, 443–456.

Schweber, S. S. (1977). The origin of the origin revisited. *Journal of the History of Biology*, *10*(2), 229–316.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, *3*, 1–5.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (Ed.). (1984). *Mechanisms of cognitive development*. New York: Freeman.

Sternberg, R. J. (1986). Cognition and instruction: Why the marriage sometimes ends in divorce. In R. F. Dillon & R. J. Sternberg (Eds.), *Cognition and instruction* (pp. 373–382). New York: Academic Press.

Sulloway, F. J. (1982). Darwin and his finches: The evolution of a legend. *Journal of the History of Biology*, *15*(1), 1–53.

Toulmin, S. (1972). *Human understanding* (Vol. 1). Princeton, NJ: Princeton University Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Cambridge, MA: Harvard University Press.

76

QUALITATIVE CHANGES IN INTUITIVE BIOLOGY

G. Hatano and K. Inagaki

Recent studies on children's intuitive biology have indicated that a form of autonomous biology is acquired early in childhood and that later qualitative changes occur within the domain. In this article we focus on two of such changes: (a) In predicting behaviors and attributing properties to an animate object, young children rely on the target's similarity to people, whereas older children and adults use its category membership and category-behavior (or property) associations; and (b) The modes of explanation change from vitalistic to mechanistic. Whereas young children prefer vitalistic explanations, older children and adults like mechanistic explanations better. We present some experimental findings for these changes. We also indicate how social contexts induce or enhance conceptual change. We discuss three theoretical issues: implications for conceptual change in biology, for conceptual change in general, and for biology instruction.

The notion of conceptual change has grown more and more popular in the area of cognitive development and in education. This idea in cognitive development was proposed (Carey, 1985) and has continued to be used (e.g., Carey, 1991) against the "enrichment" view. It posits that knowledge acquisition in such domains as naive physics, intuitive biology, and the developing theory of mind involves restructuring, i.e., that knowledge not only increases in quantity but also becomes reorganized in the course of development. At the same time, it has offered a promising alternative to Piaget. The key idea of conceptual change presupposes the domain-specificity of cognitive growth, instead of focussing on general logical structures as Piaget and his followers did. It also ascribes, unlike the Piagetian formulation, even to young children a coherent body of knowledge in those domains, which provides an explanatory framework for the target phenomena. Conceptual

Source: *European Journal of Psychology of Education*, 1997, 12(2), 111–130.

change is equated to restructuring of this body of domain-specific knowledge in which a set of concepts is embedded.

It is generally agreed that humans have a basic tendency to construct a model or theory to help make sense of an observed set of data. This implies that the initial model or theory, constructed based upon a limited database, has to be revised as more and more facts are incorporated into it, unless the initial set of observed facts constitutes a representative sample of all relevant facts. Some innate principles constrain this process of construction in a few selected domains that are critical for survival. However, they allow a variety of theories, and some of them may be weakened or given up, as accumulated pieces of prior knowledge come to serve as constraints. In this sense, conceptual change during childhood is inevitable in those domains.

Considering that a major goal of science education is to promote the construction of scientifically plausible models of the world as well as scientifically acceptable modes of prediction and causal explanation, the findings on conceptual change in cognitive development should be directly relevant. In fact, the idea of conceptual change had been discussed in science education even before Carey (e.g., Hewson, 1981), and has been relied on extensively. However, cognitive developmentalists' contributions have been rather limited; although their findings are effectively used to specify students' pre-, and to a lesser extent, post-change knowledge systems, they do not yield enough information to aid in designing instruction to induce or facilitate conceptual change in science lessons.

This limited contributions by cognitive developmentalists seem to be attributed to their investigating a different level of conceptual change from science educators. There are two conceptually distinctive levels at which the knowledge system is restructured (Vosniadou, 1994; Wellman, 1990): (1) changes in individual conceptions of what entities or phenomena are like, or in specific theories including these conceptions and (2) changes in the framework theory as a whole, including the definition of the target domain, general principles of the domain, and modes of inducing predictions and offering causal explanations. Cognitive developmentalists are primarily interested in changes in the framework theory, which are supposed to depend less on children's particular experiences. In contrast, science educators have paid more attention to changes in specific theories, which could readily be the goal of educational intervention.

However, because changes in specific theories are constrained by the framework theory even when they are data-driven and progressive (Wellman, 1990, p. 126), a change in a specific theory or its important constituent concepts is often accompanied, or even induced, by a change in the framework theory. For example, the shift from viewing a plant as an entity which takes in water to energize itself to viewing it as an entity capable of producing nutriment by itself (through photosynthesis) is apparently a change in an individual conception, but it implies a change in the ontological distinction

between animals and plants. The change from attributing diseases to moral misconducts to attributing them to such processes as contagion and contamination occurs, if it really does, at the level of a specific theory, but indicates the acquisition of an autonomous domain of biology separated from psychology. In reverse, shifts in the legitimated modes of reasoning in the framework theory, such as the change from the similarity-based to category-based inference or the change from vitalistic to mechanistic causality, could be recognized only as changes in specific concepts or theories, e.g., about attributes of a variety of animals and plants, bodily processes underlying our survival, etc. Therefore, the single notion of conceptual change can be shared by these two groups of researchers, developmental and instructional.

Therefore, contributions of cognitive-developmental research have been limited, we believe, mainly because it is yet to specify causes and consequences, as well as processes or mechanisms, of conceptual change (D. Kuhn, 1989). What is urgently needed by science educators is, among others, to better formulate the children's experiences which promote conceptual change.

The remaining part of this paper, consisting of four sections, examines in detail the shift from young children's naive biology to lay adults' intuitive biology, and tries to advance toward a goal of illuminating the notion of conceptual change for educators in the domain of biology. First, we describe strengths and weaknesses of young children's knowledge system concerning biological phenomena and processes, with the purpose of sepcifying what has to be acquired for it to become the intuitive biology that lay adults in technologically advanced societies possess. Second, we summarize our experimental findings of how these qualitative changes occur. Third, we discuss the socio-cultural contexts of these changes. Finally, we try to derive from the preceding discussions some implications for theories of conceptual change in biology and in general, and for biology instruction.

Initial system of biological knowledge

What is the initial state of conceptual change in biology in the middle and late childhood? In other words, how do we characterize young children's body of knowledge about biological phenomena? We characterize it as personifying and vitalistic in nature but as constituting a form of biology. We briefly present experimental findings to support this characterization (See Hatano & Inagaki, 1994a for details).

We assert that the body of knowledge young children possess about biological phenomena constitutes a form of biology, because it has three essential components (Inagaki, 1993b). The first element is knowledge enabling one to specify objects to which biology is applicable; in other words, knowledge about the living-nonliving distinction, and also about the mind-body distinction. The second is a mode of inference which can produce consistent and

reasonable predictions for attributes or behaviors of biological kinds. The third is a non-intentional causal explanatory framework for behaviors, properties, and bodily processes needed for individual survival and reproduction.

Living-nonliving distinction

Recent studies revealed that even preschool children can distinguish animals from nonliving things in terms of the ability to make self-initiated movements (e.g., Bullock, 1985; Massey & Gelman, 1988), possession of specific, primarily observable, properties (Gelman, Spelke, & Meck, 1983), or natural transformations over time (Rosengren, Gelman, Kalish, & McCormick, 1991). More recent studies dealing with the living-nonliving distinction including plants also indicated that young children can recognize both animals and plants as distinct from nonliving things in terms of growth (Inagaki, 1993a), regrowth by healing (Backscheider, Shatz, & Gelman, 1993), causal mechanisms involved in color transmission (Springer & Keil, 1991), and origins of object properties (Gelman & Kremer, 1991). Inagaki and Hatano (1996a) concluded, through their three studies, that 5-year-olds have a concept of living things including animals and plants differentiated from nonliving things.

Mind-body distinction

Although studies dealing with young children's ability to distinguish between the body and the mind are small in number, the available data show that young children can distinguish functions of the body from those of the mind. In other words, they differentiate biological phenomena from social or psychological ones, both of which are observed among a subset of animate things.

For example, Siegal (1988) reported that children aged 4 to 8 recognize that illness is caused not by moral but by medical factors. Experimental findings on the inheritance of biological properties indicated that at least by age six children distinguish biological parentage from adoptive parentage (Solomon, Johnson, Zaitchik, & Carey, 1996), and parentage from friendship (Springer, 1992) in attributing biological/psychological properties. Inagaki and Hatano (1993) showed that even children aged 4 and 5 have already recognized not only the differential modifiability among characteristics that are unmodifiable by any means (e.g., gender), that are bodily and modifiable by exercise or diet (e.g., running speed), and that are mental and modifiable by will or monitoring (e.g., forgetfulness), but also the independence of activities of bodily organs (e.g., heartbeat) from a person's intention. Coley (1995) revealed that kindergarten children attribute to living things biological (e.g., has blood) and psychological properties (e.g., can get angry) differently.

Mode of inference

When children do not have enough knowledge about a target animate object, they can make an educated guess by using personification or the person analogy in a constrained way. Young children are so familiar with humans that they can use their knowledge about humans as the source for analogically attributing properties to less familiar animate objects or for predicting the reactions of such objects to novel situations.

However, young children do not use knowledge about humans indiscriminately. In other words, they can use personification or the person analogy in an adaptive way in that they generate answers without committing many overpersonifying errors. In Inagaki and Hatano (1987), kindergarten children were asked to predict reactions of a rabbit, a tulip, or a stone to novel situations and explain them. These situations concerned four biological phenomena, such as too much watering and inevitable growth. Example questions are: "Suppose someone is given a baby X and wants to keep it forever in the same size, because it's so small and cute. Can he/she do that?" (inevitable growth); "Suppose X is dead tired and not lively. Will it become fine if we leave it as it is?" (spontaneous recovery).

For a rabbit, 75% of the children at least once gave personifying responses in these four questions, and for a tulip 63% did so, whereas these children gave virtually no personification for a stone. Let us give a few examples of the 5-year-olds' personifying responses. A boy answered for the inevitable growth question, "We can't keep it [a rabbit] forever in the same size. Because, like me, if I were a rabbit, I would be 5 years old and become bigger and bigger." Another boy answered for the spontaneous recovery question, "A tulip is the same as a person only on this point. If we leave it as it is, and give a little water and let it take a rest, it will become fine." It should be noted that these personifying responses tended to be associated with reasonable predictions. Inagaki and Hatano (1991) confirmed the above results through individual data analyses using a grasshopper and a tulip as targets.

Nonintentional causality

Young children cannot give articulated mechanistic explanations when asked to explain biological phenomena (e.g., bodily processes mediating input-output relations) in an open-ended interview (e.g., Gellert, 1962); sometimes they try to explain them using the language of person-intentional causality (Carey, 1985). These findings apparently support the claim that young children do not yet have biology as an autonomous domain. It seems inevitable to accept this claim so long as we assume only two types of causalities, i.e., intentional causality versus mechanistic causality, as represented by Carey (1985). However, Inagaki and Hatano (1993) propose an intermediate form of causality between these two. Children who are reluctant to rely

on intentional causality for biological phenomena, but not as yet able to use mechanistic causality, often rely on an intermediate form of causality, which might be called "vitalistic causality."

Intentional causality means that a person's intention causes the target phenomenon, whereas mechanistic causality means that physiological mechanisms cause the target phenomenon, more specifically, a specific bodily system enables a person, irrespective of his or her intention, to exchange substances with its environment or to carry them to and from bodily parts. In contrast, vitalistic causality indicates that the target phenomenon is caused by activity of an internal organ, which has "agency." The activity is often described as a transmission or exchange of the "vital force," which can be conceptualized as unspecified substance, energy, or information. Vitalistic causality is clearly different from person-intentional causality in the sense that the organ's activities inducing the phenomena are independent of the intention of the person who possesses the organ. Vitalistic explanations for biological phenomena have some common features with Keil's (1992) teleological-functional explanation for biological properties. Both are in-between the intentional and the mechanical and seem to afford valid perspectives of the biological world. However, unlike the teleological-functional explanation, vitalistic explanation is applied only to animate entities, and thus is distinctly biological.

We present below two pieces of evidence for young children's vitalism. The first was obtained from children's justifications for their predictions about bodily processes. Inagaki and Hatano (1990) reported that when asked somewhat novel questions about bodily processes, such as, "What will happen with your hands if blood doesn't come to them?" followed by, "Why do you think so?", at least one-fifth of the 6-year-olds gave "vitalistic" explanations, using expressions seemingly referring to vital power. For example, one of the children answered, "(If blood doesn't come to the hands,) they will die, because blood does not carry energies to them." Another child answered for another question "(If we don't eat food,) energies will fade away and we shall die."; "Nutriment has got out of the stomach. [What is nutriment?] It is something that gives energy."

The second piece of evidence was obtained by requiring children to make a choice from answer alternatives. Inagaki and Hatano (1993) predicted that even if young children could not apply mechanistic causality, and if they could not generate vitalistic causal explanations for themselves, they would prefer vitalistic explanations to intentional ones for bodily processes. Thus, subjects were asked to choose one from three possible explanations for each of six bodily processes, such as blood circulation, breathing, and so on. The three causal explanations represented intentional, vitalistic and mechanistic causality, respectively. An example question was: "Why do we take in air? (a) Because we want to feel good [intentional]; (b) Because our chest takes in vital power from the air [vitalistic]; (c) Because the lungs

take in oxygen and change it into useless carbon dioxide [mechanistic]. The results showed that the 6-year-olds chose vitalistic explanations as most plausible most often; they chose them 54% of the time. It should be noted that the 6-year-olds applied non-intentional (vitalistic plus mechanistic) causalities to these bodily processes 75% of the time.

A form of biology

Based on the above results, we conclude that children as young as six years of age possess three essential components of biology, and thus they have acquired a form of biology. This biology is personifying and vitalistic in nature, but is separated from psychology. Whether children of 5 years and younger possess an autonomous domain of biology is still debatable. Although preschool children seem to be able to make the ontological distinction between animate and inanimate entities (e.g., Wellman & Gelman, 1992), it is still unclear whether they understand any biology-specific causal mechanisms; studies to date provide apparently conflicting findings (e.g., Gelman & Wellman, 1991; Hirschfeld, 1995; Solomon et al., 1996).

We argue that personifying inferences are not always psychological for two reasons. First, humans are a species of living kinds, though an atypical one, and thus inferences based on knowledge about humans can be biological and even adaptive in everyday biological problem solving. As will be described in detail later, even adults rely on person analogies to some extent, as a fallback strategy, in situations where quick responding is required. Second, as described above, a considerable number of the children used personification in predicting and explaining a plant's reactions to novel situations, and these personifying explanations were reasonable from the biological point of view. It is generally agreed that plants are not included in the domain of intuitive psychology.

Vitalistic causality is different from intentional causality, which is primarily used for psychological phenomena. It should be noted that young children rely on vitalistic causality only for biological phenomena; they don't attribute social-psychological behavior, which is optional and not needed for survival, to the agency of a bodily organ or bodily part. Inagaki and Hatano (1993) found that 6-year-olds differentially chose intentional causal explanations for psychological phenomena, and vitalistic ones for biological phenomena. In other words, these children considered vitalistic causality to be used primarily to explain biological phenomena.

Young children seem to be reluctant to attribute human properties, such as "speaks to a person", to bodily organs. In addition, Hatano and Inagaki (1994b) revealed that both 6-year-olds and 5-year-olds apply organ-intentional vitalistic causal explanations much less often than organ-agential ones for biological phenomena. For example, when asked to choose one from two explanations for why we eat food everyday, "Our stomach hates it

to be empty" (organ-intentional) and "Our stomach takes vital power from food" (organ-agential), they chose the latter as more plausible much more often than the former. Hence, vitalistic causality is not intentional in the wide sense, and thus probably not psychological.

Although we will not go into detail here about how young children can acquire a form of biology so early before schooling, we would like to stress both innate and experiential bases for their acquisition. Young children seem to regard living things, based on innate domain-specific principles, as those which are similar to us humans in the sense that they take in vital force from food and/or water to maintain vigor, with its surplus inducing growth. This triangular structural relationship of "food and/or water" – "active and lively" ("becomes active by taking in vital power from food") – "grow" ("surplus vital power induces growth") is applied readily to animals. It may also be applied to plants, partly because children lack an understanding of photosynthesis. We assume that this relationship constitutes the "core" of young children's understanding of biological kinds.

At the same time, we would like to emphasize that socio-cultural constraints also play an important role in the acquisition of intuitive biology. These external constraints provide sets of specific pieces of information needed to instantiate domain-specific abstract principles of biology. For example, children's activity-based experiences contribute to this acquisition. Some such experiences are no doubt universal, but others may vary and thus produce differently instantiated versions of young children's biology. For example, when children are actively engaged in raising animals, they often acquire a richer and better structured body of knowledge about them, and thus a version of biology in which both a human and a raised animal are used as reference points (Inagaki, 1990). Larger cultural-historical contexts may also influence the acquisition of young children's biology. In fact, the biological understanding observed in different cultures is by no means identical. Even among highly industrialized countries like Israel and Japan, religious, cultural and linguistic factors produce differently instantiated versions of children's biology (Hatano, Siegler, Richards, Inagaki, Stavy, & Wax, 1993; Stavy & Wax, 1989).

From children's personifying and vitalistic biology to adults' intuitive biology

Weaknesses of personifying and vitalistic biology

In order to examine conceptual change in intuitive biology, let us start with specifying the differences between young children's biological knowledge and the intuitive biology that ordinary adults possess. In other words, we will clarify what needs to be incorporated and/or modified for young children's personifying and vitalistic biology to become lay adults' intuitive biology.

Although we have emphasized the strengths of young children's biology in the above discussion, it surely has weaknesses as well. The weaknesses are obvious even when compared with the intuitive biology in lay adults. Let us list some major ones: (a) limited factual knowledge, (b) lack of inferences based on complex, hierarchically organized biological categories, (c) lack of mechanistic causality, and (d) lack of some basic conceptual devices (e.g., "evolution," "photosynthesis").

Whereas the accumulation of more and more factual knowledge can be achieved by enrichment only, the use of inferences based on complex, hierarchically organized biological categories and of mechanistic causality requires a conceptual change (i.e., a shift in modes of reasoning). Whether the acquisition of basic conceptual devices in scientific or school biology is accompanied by conceptual change is not beyond dispute, but incorporating them meaningfully into the existing body of knowledge can usually be achieved only with the restructuring of the relevant specific theory. What follows are some experimental findings regarding these conceptual changes during childhood.

From similarity-based to category-based biological inferences

As children grow older, their personifying and vitalistic biology gradually changes toward truly "non-psychological" (if not scientific) biology by eliminating the above weaknesses (b) and (c), namely, toward a biology which relies on category-based inferences and which prefers mechanistic causal explanations and rejects intentional ones.

Carey (1985) reported results demonstrating human-centered inferences by young children, using the induction paradigm; when 4-year-olds were taught some novel properties on people, they attributed them to other animals to a greater extent than when taught on dogs. In contrast, 10-year-olds and adults who were taught on dogs were hardly distinguishable in attributional patterns from those taught on people. Rather, projections from dogs were slightly greater than those from people. These results indicate that the status of humans changes from that of a prototype to what is not more prototypical than dogs.

Inagaki and Sugiyama (1988) also examined how young children's human-centered or "similarity-based" inference would change with increasing age. They gave attribution questions about anatomical/physiological properties (e.g., a heart) to children aged 4 to 10 and college students in the form: "Does X have a property Y?" Results indicated that there was a progression from 4-year-olds' predominant reliance on similarity-based attribution (attributing human properties in proportion to perceived similarity between target objects and humans) to adults' predominant reliance on category-based attribution (attributing properties by relying on the higher-order category membership of the targets and category-attribute associations).

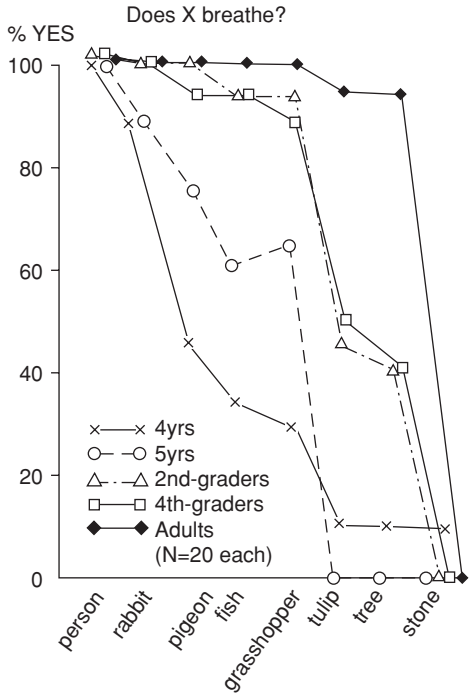


Figure 1 An example of developmental patterns obtained in attribution of anatomical/physiological properties

The intermediate pattern of attribution, which might be called "similarity-based attribution constrained by biological categories," was found between similarity-based and category-based attribution, mostly among 2nd-graders and 4th-graders. Figure 1 shows an example of developmental patterns obtained in the attribution of anatomical/physiological properties. Inagaki and Sugiyama reported that this shift was obtained through the analyses of not only group data but also of individual data. The shift seemed to occur primarily during the elementary school years.

We assume that this change is almost universal, at least among children growing up in highly technological societies. We also assume that it can occur, unlike the acquisition of basic conceptual devices, without systematic instruction in biology, though schooling may have some general facilitative effects.

From vitalistic to mechanistic causality

Young children's personifying and vitalistic biology gradually changes toward a biology which prefers mechanistic causal explanations to intentional or vitalistic ones. In Inagaki and Hatano's (1993) study described above, not

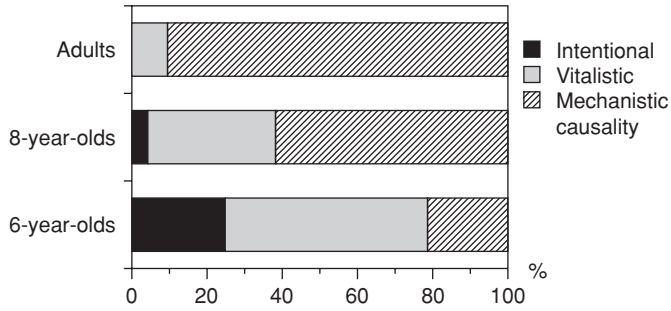


Figure 2 Percentages of choices for different types of causal explanations

only 6-year-olds but also 8-year-olds and college students were asked to choose one from three possible causal explanations for bodily processes. They found that with increasing age, the subjects came to choose mechanistic explanations as most convincing most often (see Figure 2). The 6-year-olds chose vitalistic explanations most often (54%), and intentional ones second most often. The 8-year-olds chose mechanistic explanations most often (62%), and they opted for some vitalistic ones as well (34%), but seldom chose intentional explanations. The adults predominantly preferred mechanistic explanations to explanations of the other two types. It should be noted that the difference between 8-year-olds and adults was smaller than the difference between 6-year-olds and 8-year-olds in terms of the preference of mechanistic causal explanations and the rejection of intentional ones. Individual data analyses showed that 9 out of the 20 6-year-olds chose four or more vitalistic explanations out of the six (vitalistic responders), whereas there was only one such responder among 8-year-olds. Instead, 10 out of the 20 8-year-olds were mechanistic responders (i.e., chose four or more mechanistic explanations out of the six), and 19 out of the 20 adults were so.

This result suggests that change from a reliance on vitalistic causality to a reliance on mechanistic causality occurs during the elementary school years. This change in biological causality is also assumed to be almost universal, at least among children growing up in highly technological societies, and to occur without systematic instruction in biology.

Naive biological inference as a default strategy for adults

So far, we have presented the experimental evidence suggesting that conceptual changes from naive (i.e., personifying and vitalistic) biology to adults' intuitive biology occur. Here, we would like to emphasize that despite these changes, components in the pre-change knowledge system do not disappear completely, and may be relied on as fallback strategies.

It is entirely false that adults' intuitive biology is no longer personifying at all or that their biology no longer relies on vitalistic causality. The fact that there exists a shift from similarity-based to category-based inferences does not mean that older children and adults never rely on the similarity-to-people in their attributions. Inagaki and Sugiyama (1988) reported that a substantial number of adults, as well as older children, still rely on the similarity-to-people in attributing mental properties (e.g., feeling happy) to varied animate entities. More specifically, more than 50% of adults' attribution patterns for mental properties were judged as similarity-based and only about 40% of them were category-based.

Morita, Inagaki, and Hatano (1988), using reaction time measures, revealed that to some extent college students use the similarity-to-people not only for mental properties but also for anatomical/physiological ones in situations in which they have to respond so quickly that they are not able to rely on the category membership of target objects and category-property relationships. More specifically, college students tended to make more "yes" responses to the animals more similar to people (a tortoise) than to their counterparts (a snake) within the same superordinate category (reptiles; a tortoise is regarded more similar to people than a snake is). In addition, when their responses were identical within pairs belonging to the same superordinate category but differing in rated similarity to humans, "Yes" responses were quicker to the more similar members than to the less similar ones. In contrast, "No" responses were slower to the more similar members. This interaction effect, which was statistically significant, is interpreted to mean that subjects first make a similarity-based attribution which generates a stronger tendency to respond "Yes" to the more similar members, and then they check the plausibility of this judgment, using additional knowledge including categorical knowledge. These results strongly suggest that personification or the person analogy may be used even by adults as a fallback strategy.

The fact that college students strongly preferred mechanistic to vitalistic causality (Inagaki & Hatano, 1993) does not mean that they never rely on the latter in any situation. One of the college students in their study consistently chose vitalistic explanations, and in the interview after the experiment said, "We usually choose something like 'oxygen' or 'the heart works like a pump' because we have learned in school to do so. However, I chose other explanations because they were most convincing and comprehensible to me." This suggests that vitalistic causality continues to work as a basis of understanding and to be used in situations where people do not think they are required to provide answers based on so-called scientific biology. It also suggests that conceptual change toward an exclusive reliance on mechanistic causality is, at least in part, due to social pressure, and is thereby not a purely cognitive process.

Incorporating scientific biological concepts

The fourth and last weakness of young children's naive biology mentioned above is the lack of basic conceptual devices of scientific biology. In other words, young children are not capable of viewing the biological world with scientific conceptual tools. In order to be able to understand and reason "scientifically" in biology one needs to know its basic concepts and principles as major conceptual devices. For example, one who does not know the phenomenon of photosynthesis cannot understand the basic difference between animals and plants (i.e., plants can produce nutriment themselves), and thus may accept the false analogy of mapping water for plants to food for animals. We assume that, unlike the conceptual change in inference and causality described above, this change is very hard to bring about, especially without educational intervention, and thus occurs only among a limited portion of older children or adolescents.

How do children incorporate scientific concepts and change their knowledge system into a more advanced one? Let us take the case of evolution. The (Darwinian) idea of evolution must be difficult for children to grasp. It has been fully accepted even among biologists only within the last century. However, because naive biology assumes living things, but not nonliving things, to be able to adjust themselves to their ecological niche or ways of life, children are ready to accept any biological kind's gradual adaptive changes over generations, and thus to form a version of the Lamarckian idea of evolution (Marton, 1989). We will describe in the next section how elementary school children's understanding of evolution can be modified, specifically, how a change in specific biological theory can take place through whole class discussion.

It should be noted that, as children learn these and other scientific conceptions in school biology, their ways of understanding the biological world may in fact change. In other words, not only is school biology learned meaningfully by being assimilated into existing knowledge of naive biology, but also, as claimed by Vygotsky (1978), it reorganizes naive biology by adding, say, physiological mechanisms and the evolutional perspective, so that the reorganized body of knowledge can effectively be used as the basis for answering a wider variety of biological questions.

Socio-cultural contexts of conceptual change

Young children's personifying and vitalistic biology is certainly limited primarily because their database is limited. They have to rely on the person analogy, because they do not know much about other living kinds. They have to explain biological phenomena in terms of vitalistic causality, because they are ignorant of detailed physiological mechanisms. Thus an increased amount of knowledge regarding biological processes and kinds is a necessary

condition for the appearance of a more sophisticated knowledge system, as is often assumed by conceptual change researchers. However, we would claim that the accumulation of biological knowledge is not enough for inducing conceptual change, and that the social contexts in which children are exposed to biological information are critical. For example, if children engage in activities that provide meaningful and socioemotionally-laden contexts, they are likely to acquire an advanced biological knowledge system more readily. We cite a few such studies below.

Cognitive consequences of animal-raising activities

An animal-raising activity in which children engage in meaningful contexts tends to produce a more advanced mode of inference. Inagaki (1990) compared the biological knowledge of kindergartners who had actively engaged in raising goldfish for an extended period at home with that of children of the same age who had never raised any animal. Although these two groups of children did not differ in factual knowledge about typical mammals, the goldfish-raisers had much richer procedural, factual, and conceptual knowledge about goldfish. The goldfish-raisers used their knowledge about goldfish as a source for analogies in predicting reactions of an unfamiliar "aquatic" animal (i.e., a frog), one that they had never raised, and produced reasonable predictions with some explanations for them. One goldfish-raiser, when asked whether we could keep a baby frog in the same size forever, replied, "No, we can't, because a frog will grow bigger as goldfish grew bigger. My goldfish were small before, but now they are big." These goldfish-raisers tended to use person analogies for for a frog as well, and thus they could use two sources for making analogical predictions.

In another study, Inagaki (1996; see also Hatano & Inagaki, 1992) asked both goldfish-raisers and non-raisers to attribute biological properties to a variety of animals. A plant and a stone were included as controls. Interestingly, in this context, results indicated that goldfish-raisers attributed animal properties which are shared by humans (e.g., having a heart, excreting) not only to goldfish but also to a majority of animals phylogenetically between humans and goldfish at a higher rate than non-raisers. In other words, while attributional patterns of the non-raisers were judged to be similarity-based, those of the goldfish-raisers were almost category-based. The mere presence of goldfish at home did not induce any cognitive consequences; those having goldfish at home but not taking care of them were not markedly different from the non-raisers. These results suggest that the experience of raising goldfish in meaningful contexts, probably accompanied by the desire to understand the pets and to offer a better care for them, modifies young children's preferred mode of biological inferences. In other words, it at least enhances a conceptual change in biology.

Socialization for trusting particular modes of inference and explanation

Hatano and Inagaki (1991b) examined whether the shift from similarity-based to category-based inferences would be induced, at least in part, by a metacognitive belief about the usefulness of higher-order categories, i.e., belief that the category-based inference (e.g., "The grasshopper is an invertebrate, so it must have no bones") is more dependable than the inference relying on similarity metrics (e.g., "The grasshopper is not very similar to a human, so it will have no bones"). The 2nd-, 4th- and 6th-graders were required in a questionnaire format to evaluate a given set of reasons, which were allegedly offered by the same age children in a dialogue with the teacher. That is, they were asked to judge the plausibility of three different types of reasons, each of which was preceded by a Yes-No judgment to such a question as "Does an eel have bones?" or "Does a tiger have a kidney(s)?". Two of the three reasons represented the similarity-based inference and the category-based one. The former referred to the target's surface similarity to people, such as, "I think a tiger has a kidney, because it is generally like a human", and the latter referred to higher-order categories like "mammals", such as "I think a tiger has a kidney, because both a human and a tiger are mammals." The other reason, clearly not category-based nor similarity-based, was a distracter, e.g., "I think a tiger does not have a kidney, because it is not as intelligent as a human."

Results indicated that as children grew older, the number of respondents who judged the category-based reason to be plausible and the similarity-based one to be implausible significantly increased, whereas the number of respondents judging the similarity-based reason plausible and the category-based one implausible decreased, suggesting that children came to acquire the metacognitive belief about the usefulness of higher-order categories. Moreover, even within the 2nd-graders, those children who consistently favored category-based reasons tended to show an attributional pattern closer to the pure category-based attribution than was shown by those who favored similarity-based reasons.

Hatano and Inagaki further examined, by using indirect, "projective" type questions, whether older students differentiated more clearly those fictitious children who gave a category-based reasons from those who gave a similarity-based reasons in the rating of their academic talent. Another group of 2nd-, 4th- and 6th-graders were given a questionnaire. It described two hypothetical pairs of children of the same grade as the students, who, in dialogue with their teacher, gave a judgment of whether rabbits and ants had a pancreas (or tigers and grasshoppers had bones) and the reason for it. The reason given by one of each pair was in fact similarity-based and the other, category-based (these labels were not given). The subjects were asked to rate how good academically the fictitious child who had given the reason would be, and how likable the child would be as a friend, in a four-point scale.

Results were as follows: The fictitious child who had allegedly given a category-based reason was rated significantly higher in academic talent than the allegedly similarity-based child in all grades. However, the older the subjects, the bigger was the magnitude of the difference. That is, the older children were much less positive than the younger ones in the rating of the fictitious child who had given the similarity-based reasons. Since the likability rating for these fictitious children did not differ significantly, it is not likely that these subjects always gave favorable ratings for the category-based child. These results strongly suggest that children become reluctant to use the similarity-to-people as an inferential cue for biological characteristics as they grow older. It is likely that conceptual change in modes of biological inference is enhanced by social sanctions. For instance, children may stop relying on similarity to people in order to avoid being regarded as less talented.

Conceptual change induced by classroom discussion

A joint attempt at comprehension by a group often leads to the acquisition of more sophisticated knowledge than that which would be obtained by independent efforts by its members. Comprehension activity, whether individual or collective, includes proposing a possible interpretation, offering evidence for or against the interpretation, deriving a prediction from the interpretation, testing the prediction, evaluating the tested result, proposing another interpretation, and so on. Because the group as a whole has a richer database than any of its members, it is likely that more varied interpretations can be offered, and that denser and less biased pieces of evidence can be presented jointly than individually. Thus, conceptual change may also be induced through discursive interactions.

Hatano and Inagaki (1991b) examined, using a whole-class discussion called the Hypothesis-Experiment-Instruction (HEI) method, whether each of students who engaged in collective comprehension activity is likely to acquire more elaborate knowledge through sharing understanding than those who did not. The HEI method entails the following steps: predictions of a scientific phenomenon, collective examination of reasons for the predictions through discussion, and confirmation of the predictions. In this study, the issue of evolution, specifically, the characteristics of monkeys in relation to their lives in trees, was used as the target subject.

Experimental groups of about 20 fifth-graders each first read a short passage about the relationships between animals' characteristics and their ways of living. Next, they were given a problem in multiple choice form about the monkeys' characteristics. The target problem was: "Do the thumbs of monkeys' forefeet oppose the other 'fingers' (like in human hands) or extend in parallel to other 'fingers' (like toes in human feet)? How about the thumbs of their hind feet?" Answer alternatives included: (a) The thumbs

are never opposing, (b) The thumbs are opposing only in the forefeet, (c) The thumbs are opposing in both fore and hind feet. [Alternative (c) is correct.] Pupils' response frequencies were tabulated on the blackboard, and group discussion followed. After about 15 minutes' discussion, the pupils chose an answer alternative once again. Finally, they were given a short passage stating the correct answer. All that this passage described was the relevant facts about monkeys. In other words, it contained no explanations about why these characteristics had evolved.

A control condition was provided to assess how likely it was for such pupils to construct knowledge without social interaction. Control groups of about 20 pupils each were given the same passage to read immediately after answering, for the first and only time, the multiple choice problem. After reading the passage about the characteristics of monkeys, both experimental and control pupils were (individually) asked to explain why monkeys had a thumb opposing their other fingers.

Results revealed that the pupils in the experimental group gave significantly more elaborate explanations than did the control subjects, by connecting the given facts in the passage to some of the ideas expressed in the discussion. This was true even for those pupils who had never expressed their opinion during the discussion. Let us show two examples of these "silent" experimental subjects.

M.Y. (a boy) chose the alternative of (b) before the discussion. When asked, in the questionnaire after the discussion, to give names of peers who gave reasonable opinions during the discussion, he referred to two "vocal" supporters of (b) as those whose opinions had been the same as his. At the same time he named a girl who had supported (c) as a proponent of a reasonable explanation. He did not change his prediction, and his curiosity and confidence after the discussion increased. His explanation on the posttest was rated elaborate: "A monkey cannot climb a tree nor grasp an object unless its fore and hind feet are thumb-opposing." This suggests that he incorporated information from the supporter of (c) when he read the material and found out that his idea was correct for the forefeet but not for the hind feet.

T.I. (a girl) did not try to find an agent in discussion; she did not name anybody whose opinion had been plausible. She chose alternative (c) both before and after the discussion, and wrote in the questionnaire that she had not changed her answer because she had been confident in herself. Her explanation on the posttest was, "Because [monkeys] have not walked on the ground often, [their feet] have become suitable for holding on to branches." Since no control group pupils gave such explanations, we can infer that she was responding to the argument by supporters of (b), comprising the majority, that "thumb-opposing hind feet are inconvenient for walking."

Needless to say, what were observed in this study were mostly changes in individual pieces of knowledge or beliefs, not a change in the conception of evolution. Achieving joint comprehension does not guarantee the occurrence

of conceptual change. However, the study at least suggests that whether one's own beliefs are maintained, elaborated, or discarded depends on the nature of the collective comprehension activity, particularly the dialogical processes and dynamics regarding the target of comprehension. Here a variety of socio-emotional factors (e.g., likes and dislikes of the proponent and/or the opponent, motivation to form a majority, etc.) as well as cognitive factors are involved.

Moreover, some of the belief changes were accompanied by more profound restructuring of the knowledge system. For example, several pupils changed their ways of reasoning about a monkey's physical characteristics. Initially basing their reasoning on the animal's similarity to humans, they ended up taking the adaptation to its ways of living as critical. This could be a step toward the acquisition of a more sophisticated (Lamarckian) conception of evolution with accompanying changes in related beliefs about the specific theory of biological taxonomy.

Theoretical and educational implication

Recent studies on children's intuitive biology have indicated that a form of autonomous biology is acquired early in childhood and that conceptual changes within the domain occur later. We believe that such a conceptual change is necessary, because enrichment is sufficient only when innate constraints continue to be very strong and initial observations provide an unbiased sample. These conditions may be met for the basic taxonomy (e.g., classifying things into animals, plants, and inanimate things): Atran (1994) reports that there exists a cross-cultural universality in aspects of folk taxonomy of animals and plants. However, the apt attribution of unobservable anatomical/physiological properties to unfamliar entities, the differentiation between the mental and the bodily, and the causal explanation for the bodily processes require qualitative changes: changes toward abandoning those initally effective modes of reasoning, changes toward constructing a non-human-centered, hierarchically organized classification system, changes toward offering more specified causal mechanisms, and so on.

In this article we have focussed on two such qualitative changes. One is the change in modes of inference: when predicting behaviors of or attributing properties to an unfamiliar animate object, young children rely on similarity-based inference, whereas older children and adults use category-based inference. The other is the change in modes of explanation, from vitalistic to mechanistic.

We believe that, as the framework theory of biology changes, there occur, with some horizontal decalage, corresponding changes in specific theories regarding nutrition and growth, bodily processes, diseases, biological parentage, evolution, and so on. Although more local changes in one of these specific theories may be induced in the data-driven fashion, the depersonalizing

and devitalisting change takes place everywhere in the domain of intutive biology.

In this final section, we discuss three issues: (a) how the above conclusions about conceptual change in children's intuitive biology are compatible with other, major characterizations of it, (b) what these conclusions offer to our understanding of conceptual change in general, and (c) instructional implications of these conclusions.

Comparison with other models of conceptual change in biology

There is a consensus among recent developmentalists that the acquisition of biological knowledge undergoes conceptual changes in childhood. More specifically, many developmentalists acknowledge that young children's biological knowledge system is not the same as the intuitive biology that older children and lay adults have, and thus that it undergoes qualitative changes in childhood. For example, Carey and Spelke (1994) state that "it is clear that their [preschool children's] understanding of biological phenomena differs radically from that of older children," and that "children progress from vitalistic biology to mechanistic biology" (p. 185). Even Keil, who seems to be on the side of the enrichment view, states in his recent paper (Keil, 1994) that "a great deal of conceptual change does occur with respect to biological thought in the first 10 years of life" (p. 250).

However, there are some disagreements among these developmentalists concerning the nature of the initial system of biological knowledge, and whether conceptual change occurs across domains or within a domain (Although there is a debate concerning when the initial biology is acquired, at preschool age or at 6 or 7 years of age, we will not deal with this issue separately here). Carey (1985) claimed that an intuitive biology emerges from an intuitive psychology between ages 4 and 10, and that preschoolers' initial system of knowledge about biological phenomena does not yet constitute an autonomous domain of biology. In other words, she proposed that conceptual change in the biological knowledge system occurs across domains, i.e., in the form of the differentiation of biology from the domain of psychology. Based on the findings of many empirical studies done after Carey (1985), she recently modified her original claim concerning when intuitive biology is acquired (Carey 1995) from age 10 to an earlier age, say, age 6 or 7. However, she still holds the position that an intuitive biology emerges from an intuitive psychology (Carey, 1995), in other words, conceptual change across domains takes place at an earlier age, say ages 3–6.

In contrast, Keil (1992, 1994) claims that young children's initial system of biological knowledge is never psychologically driven; rather, it constitutes a distinct biological theory or mode of construel from the beginning, and thus that even children younger than 6 years of age possess an autonomous domain of biology. Although he acknowledges in his recent paper (Keil,

1994) that conceptual change occurs within the domain of intuitive biology, as described above, he does not specify at all what kind of change it is.

We agree with Keil (1992, 1994) and Wellman and Gelman (1992) that intuitive (or naive) biology is a core domain, like intuitive psychology and physics, which is more or less innately constrained, and that domains of biology and psychology are separated from each other among children younger than 6 years of age. We speculate, from the perspective of human evolution, that innate constraints help us establish the domain of biology, because it has been vital for our species to have some knowledge about animals and plants as potential foods (Wellman & Gelman, 1992), and also knowledge about our bodily functions and health (Inagaki & Hatano, 1993).

However, contrary to Keil (1992, 1994), we do not view young children's initial system of biological knowledge as completely free from psychological influences. We speculate that young children's biology is acquired a little later than their intuitive physics or intuitive psychology. More specifically, it is gradually constructed, based on skeletal guiding principles unique to the domain of biology, through daily experience in early years. Thus, the construction of the initial biology can be affected by previously acquired systems of knowledge including the knowledge of how the mind works, that is, psychology. It is at least possible that very young children are sometimes tempted to interpret biological phenomena by borrowing psychological knowledge, because their biological knowledge is not powerful enough to generate convincing predictions and explanations by itself. In fact, as demonstrated in the previous section, young children's biological knowledge system is personifying and vitalistic in nature, in other words, it has a psychological flavor. Our more recent studies also indicate that among preschool children younger than 6 years of age a domain of biology is separate from that of psychology, but their biology is much influenced by psychology (Inagaki & Hatano, 1996b).

Anyway, we believe that the biology children have acquired at least by age 6 qualitatively changes in the middle and late childhood from personifying and vitalistic toward more complexly taxonomic and mechanistic within the autonomous domain of intuitive biology. However, unlike Carey's earlier claim (1985), we argue that this personifying and vitalistic knowledge system constitutes a form of biology, already separated from intuitive psychology, and, as supported by recent Carey's claim (1995), this personifying and vitalistic biology undergoes conceptual change within the domain.

Implications for the notion of conceptual change

In addition to contributing to the understanding of conceptual change in intuitive biology, the research findings reviewed above may have some implications for the notion of conceptual change in general, more specifically, its consequences, causes, and processes. Let us mention them briefly in this order.

First, the reviewed studies indicate that, though the pre- and post-change knowledge systems are qualitatively different, there are some continuities between them. The occurrence of restructuring does not mean that components of the pre-change system disappear completely in the post-change system. These components may not be discarded. Considering recent research findings revealing multiple models (Yates, Bassman, Dunne, Jertson, Sly, & Wendelboe, 1988) and multiple strategies (e.g, Siegler & Jenkins, 1989) within the same subjects, the most likely consequences is that old components stay as less salient fallback models or strategies in the new system. An important implication of this conclusion is that the postchange knowledge system of educated adults may not be as drastically different from young children's pre-change knowledge system as it appears.

Second, the studies show that conceptual changes are often induced by participation in goal-directed activities, and enhanced by discursive processes in a group. We fully agree that the increased amount of knowledge is the necessary condition for conceptual change (e.g., Carey, 1985; Smith, Carey, & Wiser, 1985; Wiser, 1988) and that the pre-change system serves as cognitive or internal constraints in conceptual change. However, we would also like to emphasize the role of other people and tools as socio-cultural or external constraints in conceptual change. We are afraid that most leading investigators studying conceptual change have been too cognitive and too individualistic. The issue of motivation inducing conceptual change rather than local patchwork has generally been neglected, with a few notable exceptions (e.g., Pintrich, Marx, & Boyle, 1993).

Finally, the reviewed studies suggest two mechanisms for conceptual change or the process of restructuring. One is the spreading of the truth-value alteration, which can be described by expanding "symbolic connectionist" models (Holyoak, 1991). Some pieces of knowledge may be given direct feedback that changes their truth value, but the alteration of the truth value of others is induced by the change of the truth value of connected pieces (e.g., If that a goldfish excretes is true, that a frog excretes becomes more likely) or the change in the strength of connections (e.g., That humans feel sad does not imply that a grasshopper feels sad). When changes in the truth value of some pieces are accumulated, there can be a drastic change in almost all pieces through continued spreading and recurring effects. The other mechanism involves conceptual-procedural correspondence plus sociocultural sanctions to use some particular modes of reasoning or strategies. Although procedures are not fully governed by conceptual understanding (e.g., Resnick, 1982), those procedures newly chosen or discovered tend to be compatible with the corresponding conceptual knowledge. Thus, for instance, change in the taxonomy will lead to strategies exploiting taxonomy (e.g., relying on a prototype of the category to which the target belongs instead of personification after a complex hierarchy of classifications is established).

Instructional implications

The growing body of research on naive and intuitive biology, especially that on conceptual change, has significant implications for education in general, and for the teaching-learning of biology in particular. As aptly pointed out by Olson and Torrance (1996), "any attempt of teaching is premised on an understanding of the mind of the learner," and vice versa. Our understanding of how children's mind works and grows can direct, or even specify, contents and methods of teaching.

The conclusion from our review is that young children before schooling have acquired a form of autonomous biology and that this biology undergoes conceptual changes within the domain during childhood. Detailed descriptions of the initial state would be very helpful for designing an effective course of instruction, because the instruction aims at changing learner's knowledge from the initial state to the goal state (Glaser & Bassok, 1989). This conclusion implies that starting instruction of biology at kindergarten or in the lower elementary grades is possible and can be effective, but the instruction must enhance restructuring of it. This implication is clearly distinct from that of Carey (1985) and others who assume that young children have no form of biology, because, according to the latter, we have to teach biology as a totally new discipline in school or postpone its teaching until the 5th grade or so. However, it is also distinctive from that of researchers who assume only enrichment in intuitive biology, because, according to the above conclusion of undergoing conceptual change, one cannot aim for linear progress in students' biological understandingt.

The above specifications of consequences, causes, and processes of conceptual change are also suggestive. For example, because conceptual change is often induced by participation in goal-directed activities and enhanced by discursive processes in a group, educators should try to organize such social activities and interactions. This implication is clearly different from the one derived from purely cognitive and individualistic formulations of conceptual change.

As we increase our understanding of the processes of conceptual change in biology, we can better design and implement biology lessons (Inagaki, 1994). Even now, we are sure that educators should activate students' informal (pre-change) biological knowledge and relate formal biology instruction to it as much as possible, and give students an ample amount of time to incorporate new pieces of information into the existing body of knowledge. It is expected that naive biology, which is personifying and vitalistic in nature, can provide students with a conceptual framework for learning school biology meaningfully. At the same time, the naive biology can become a more mature version of intuitive biology by incorporating pieces of scientific or school biology, and by spreading and recurring their influences.

References

Atran, S. (1994). Core domains versus scientific theories: Evidence from systematics and Itza-Maya folkbiology. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 316–340). Cambridge, UK: Cambridge University Press.

Backscheider, A.G., Shatz, M., & Gelman, S.A. (1993). Preschoolers' ability to distinguish living kinds as a function of regrowth. *Child Development*, *64*, 1242–1257.

Bullock, M. (1985). Animism in childhood thinking: A new look at an old question. *Developmental Psychology*, *21*, 217–225.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 257–291). Hillsdale, NJ: Erlbaum.

Carey, S. (1995). On the origin of causal understanding. In D. Sperber, D. Premack, & A.J. Premack (Eds.), *Causal cognition* (pp. 268–302). Oxford: Clarendon Press.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169–200). Cambridge: Cambridge University Press.

Coley, J.D. (1995). Emerging differentiation of folkbiology and folkpsychology: Attributions of biological and psychological properties to living things. *Child Development*, *66*, 1856–1874.

Gellert, E. (1962). Children's conceptions of the content and functions of the human body. *Genetic Psychology Monographs*, *65*, 291–411.

Gelman, R., Spelke, E., & Meck, E. (1983). What preschoolers know about animate and inanimate objects. In D. Rogers & J.A. Sloboda (Eds.), *The acquisition of symbolic skills* (pp. 297–326). New York: Plenum.

Gelman, S.A., & Kremer, K.E. (1991). Understanding natural cause: Children's explanations of how objects and their properties originate. *Child Development*, *62*, 396–414.

Gelman, S.A., & Wellman, H.M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition*, *38*, 213–244.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, *40*, 631–666.

Hatano, G. (1994). Introduction. *Human Development*, *37*, 189–197.

Hatano, G., & Inagaki, K. (1991a). Sharing cognition through collective comprehension activity. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 331–348). Washington, DC: American Psychological Association.

Hatano, G., & Inagaki, K. (1991b). *Learning to trust higher-order categories in biology instruction*. Paper presented at the meeting of the American Educational Research Association, Chicago.

Hatano, G., & Inagaki, K. (1992). Desituating cognition through the construction of conceptual knowledge. In P. Light & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 115–133). London: Harvester/Wheatsheaf.

Hatano, G., & Inagaki, K. (1994a). Young children's naive theory of biology. *Cognition*, *50*, 171–188.

Hatano, G., & Inagaki, K. (1994b). *Bodily organ's "intention" in vitalistic causal explanations*. Paper presented at the 36th annual meeting of Japanese Educational Psychology Association [in Japanese].

Hatano, G., Siegler, R.S., Richards, D.D., Inagaki, K., Stavy, R., & Wax, N. (1993). The development of biological knowledge: A multi-national study. *Cognitive Development*, *8*, 47–62.

Hewson, P.W. (1981). A conceptual change approach to learning science. *European Journal of Science Education*, *3*, 383–396.

Hirschfeld, L.A. (1995). Do children have a theory of race? *Cognition*, *54*, 209–252.

Holyoak, K.J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 301–335). Cambridge: Cambridge University Press.

Inagaki, K. (1990). The effects of raising animals on children's biological knowledge. *British Journal of Developmental Psychology*, *8*, 119–129.

Inagaki, K. (1993a). *Young children's differentiation of plants from nonliving things in terms of growth*. Paper presented at the biennial meeting of Society for Research in Child Development, New Orleans.

Inagaki, K. (1993b). *The Nature of Young Children's Naive Biology*. Paper presented at the 12th meeting of International Society for the Study of Behavioral Development, Recife, Brazil.

Inagaki, K. (1994). *Personifying and Vitalistic Biology: Its Nature and Instructional Implications*. Paper presented at the 13th meeting of the International Society for the Study of Behavioral Development, Amsterdam.

Inagaki, K. (1996). *Effects of raising goldfish on young children's grasp of common characteristics of animals*. Paper to be presented at the 26th International Congress of Psychology, Montreal.

Inagaki, K., & Hatano, G. (1987). Young children's spontaneous personification as analogy. *Child Development*, *58*, 1013–1020.

Inagaki, K., & Hatano, G. (1990). *Development of explanations for bodily functions*. Paper presented at the 32nd annual convention of the Japanese Association of Educational Psychology, Osaka [in Japanese].

Inagaki, K., & Hatano, G. (1991). Constrained person analogy in young children's biological inference. *Cognitive Development*, *6*, 219–231.

Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. *Child Development*, *64*, 1534–1549.

Inagaki, K., & Hatano, G. (1996a). Young children's recognition of commonalities between animals and plants. *Child Development*, *67*, 2823–2840.

Inagaki, K., & Hatano, G. (1996b). *Emerging Distinction between Naive Biology and Naive Psychology*. Paper presented at the XIVth Biennial Meetings of International Society for the Study of Behavioural Development, Quebec City.

Inagaki, K., & Sugiyama, K. (1988). Attributing human characteristics: Developmental changes in over- and underattribution. *Cognitive Development*, *3*, 55–70.

Keil, F.C. (1992). The origins of an autonomous biology. In M.R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition; The Minnesota Symposia on Child Psychology* (vol. 25, pp. 103–137). Hillsdale, NJ: Erlbaum.

Keil, F.C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind* (pp. 234–254). Cambridge, MA: Cambridge University Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689.

Marton, F. (1989). Towards a pedagogy of content. *Educational Psychologist*, *24*, 1–23.

Massey, C.M., & Gelman, R. (1988). Preschooler's ability to decide whether a photographed unfamiliar object can move itself. *Developmental Psychology*, *24*, 307–317.

Morita, E., Inagaki, K., & Hatano, G. (1988). *The development of biological inferences: Analyses of RTs in children's attribution of human properties*. Paper presented at the 30th annual convention of the Japanese Association of Educational Psychology, Naruto [in Japanese].

Olson, D.R., & Torrance, N. (1996). *Handbook of education and human development: New models of learning, teaching and schooling*. Cambridge: Blackwell.

Pintich, P.R., Marx, R.W., & Boyle, R.A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*, 167–199.

Resnick, L.B. (1982). Syntax and semantics in learning to subtract. In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 136–155). Hillsdale, NJ: Erlbaum.

Rosengren, K.S., Gelman, S.A., Kalish, C.W., & McCormick, M. (1991). As time goes by: Children's early understanding of growth. *Child Development*, *62*, 1302–1320.

Siegal, M. (1988). Children's knowledge of contagion and contamination as causes of illness. *Child Development*, *59*, 1353–1359.

Siegler, R.S., & Jenkins, E. (1989). How children discover new strategies. Hillsdale, NJ: Erlbaum.

Smith, Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concept of size, weight, and density. *Cognition*, *21*, 177–237.

Solomon, G., Johnson, S., Zaitchik, D., & Carey, S. (1996). Like father, like son: Young children's understanding of how and why offspring resemble their parents. *Child Development*, *67*, 151–171.

Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development*, *63*, 950–959.

Springer, K., & Keil, F.C. (1991). Early differentiation of causal mechanisms appropriate to biological and nonbiological kinds. *Child Development*, *62*, 767–781.

Stavy, R., & Wax, N. (1989). Children's conceptions of plants as living things. *Human Development*, *32*, 88–94.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*, 45–69.

Vygotsky, L.S. (1978). *Mind in society*. Edited and translated by M. Cole, S. Scribner, V. John-Steiner, & E. Souberman (Eds.), Cambridge: Harvard University Press.

Wellman, H.M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H.M., & Gelman, S.A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–375.

Wiser, M. (1988). The differentiation of heat and temperature: history of science and novice-expert shift. In S. Strauss (Ed.), *Ontogeny, phylogeny, and historical development* (pp. 28–48). Norwood, NJ: Ablex.

Yates, J., Bassman, M., Dunne, M., Jertson, D., Sly, K., & Wendelboe, B. (1988). Are conceptions of motion based on a naive theory or on prototypes? *Cognition*, *29*, 251–275.

77

DEVELOPING UNDERSTANDING WHILE WRITING ESSAYS IN HISTORY

J. F. Voss and J. Wiley

Using the textbase-situation model of discourse processing and assuming a distinction of learning (recall of text contents) and understanding (relating different parts of text contents or text to non-text contents), it was found that individuals reading text contents from a number of sources who wrote an argumentative essay about the contents and then rated content elements for importance developed a better understanding of the contents than individuals writing a narrative essay and making importance ratings either before or after writing.

The decades of the 1980s and 1990s have been marked by the study of learning and reasoning in various subject matter domains (Voss, Wiley, & Carretero, 1995). While such domains include non-school subjects such as chess, much of the research has been focussed upon the subject matter of the school curriculum, especially physics and mathematics. More recently, however, the topic of history has received increasing attention (Carretero & Voss, 1994), this chapter summarizes the results of a study conducted on this topic.

The study of subject matter learning and reasoning has in part been motivated by the desire to develop a greater understanding of how people learn, reason, and think. But the motivation has also been pedagogical; there is the desire to improve classroom instruction by learning more about how students think in subject matter terms. This motive is quite strong in the United States because national and international studies suggest American students are not acquiring appropriate knowledge and skills. In the field of history, published reports (e.g., Beatty, Reese, Persky, & Carr, 1996; Ravitch & Finn, 1987) suggest student knowledge of history generally tends to be poor.

Source: *International Journal of Educational Research*, 1997, 27(3), 255–265.

Cognitive framework

It is assumed that learning is positively related to the level of processing that occurs when an individual relates new input to his or her pre-existing knowledge. Thus, because experts are better able to integrate new information with well-developed knowledge of the subject matter, experts tend to retrieve information better than novices. The Kintsch (1994) conceptual-integration model captures this idea by postulating the existence of a situation model, which involves the activation of information in memory by the input and the integration of input contents with the activated contents of memory.

Often, however, as when reading a novel, a person may have the general characteristics of a model in memory and some knowledge of the time and location at which the story takes place, but the person, during the course of reading the novel, needs to construct a model of the plot and the characters of the novel, modifying the model as the reading continues. To include in the model all of the information in the novel is futile, in part because of limitations of working memory. As a result, the individual usually selects information in order to maintain coherence, establish causality, maintain impressions of the characters, and perform other functions.

How, then, is this process different from the process used when a student reads a history text? Based upon the results reported in this chapter and those of another study reported elsewhere (Wiley & Voss, 1996), two differences can be noted. One is that history text can occur in different forms and processing may vary with the nature of the presentation. Specifically, the material may be presented in a textbook, in a volume on a given historical topic, or in a history journal. Historical information also can be presented via records, newspaper articles and editorials, paintings, photos, and other artifacts (collectively known as sources). A question of interest then is how the nature of the source influences the processing that occurs. In this case the "history" needs to be constructed from the historical information, with such information requiring selection and integration.

A second way in which the study of history is different from reading a novel is that, in the school context, the reading of history is typically followed by some type of questioning of the student about the contents of the history assignment. The task can take the form of a multiple-choice test, the writing of an essay about a topic discussed in the assignment, the defending of a particular position about the interpretation of the historical content, or some other procedure involving performance assessment.

The above considerations suggest a distinction between the concept of learning and that of understanding. For our purposes learning will refer to a person's ability to perform on a task that measures the acquisition of some content. Thus, what and how much one remembers from the contents of a history chapter would define what the person has learned. However, understanding is taken to refer to the knowledge a person has about the underlying

conceptual relations of a given topic, the relations often including the interpretation of the presented material. A person therefore may learn quite a bit in terms of recall, but have a poor understanding, in terms of understanding important conceptual relationships related to the material. Moreover, it is assumed that learning can take place with a relatively low amount of processing whereas understanding generally involves more extensive processing, with the latter also involving greater integration of new and old information.

Wiley and Voss (1996) conducted a study in which essentially the same information was presented in two different formats. Students were asked to read material about the Irish potato famine of the mid 19th century, with the material presented either in a standard textbook format or as sources. The material included charts and graphs of population size and immigration figures, as well as social, political, and historical information. It was hypothesized that learning from sources would yield higher performance than learning from a textbook because presumably more processing would be required to integrate the source material than to integrate the material in the textbook since the textbook information was already organized. The source information, however, required conceptual integration and this required processing.

A second variable manipulated was the assigned task. The students were asked to write an essay, with each of three groups writing a different type of essay. All three groups were asked to indicate in their essay what produced the significant changes in Ireland's population between 1845–1850, with one group being asked to write a narrative, another being asked to write a history, and one group being asked to write an argument of why this occurred. The hypothesis was that writing an argument would require the most processing because that task required examining possible factors contributing to the population changes and organizing them into a reasonable argument. Writing a narrative was exected to involve less processing, while writing a history would depend upon the writer's idea of what a history is.

While the test hypotheses pertained to the main effects of the two variables under study, the primary focus was upon their interaction. It was hypothesized that the deepest or most extensive processing would occur in the condition in which students read the sources and wrote an argument. The effects of the two variables would sum, thereby producing the most extensive processing. Writing a narrative essay after reading the materials in textbook-like form was also expected to fit well together since the material and essay organizations were highly similar. However, because they mapped so well onto each other, processing would not be extensive. Thus, it was hypothesized that while this condition would produce substantial learning in terms of recall of presented information, the processing would result in limited understanding. In other words, while both textbook-like presentation of material with narrative essay condition and the source presentation with argument writing would both produce good recall of information, only the latter condition would produce high levels of understanding. Again, the

results involving the history essay condition was expected to depend upon the students' concept of a history text.

The results generally supported the hypotheses. While the textbook presentation/narrative essay condition and the source presentation/argument essay condition produced better recall than the other four conditions and recall did not vary between the two conditions, the source presentation/argument essay condition yielded superior performance on measures of understanding such as the number of connections made between textual factors and the number of causal links stated in the essay.

Present experiment

The experiment reported here is concerned with how individuals select information from sources and use it to write an essay. The procedure used in the present study was to ask the students to read information about the Potato Famine and write an essay of a particular type. When writing the essay the students were allowed to view all presented information, primarily because we wanted the focus of the work to be on essay writing, not memory. Because the students did not know the essay condition to which they were to be assigned as they read the presented information, there should be no differences during initial reading in relation to the essay condition manipulation. Given the results of Wiley and Voss (1996), it was hypothesized that individuals in the argument essay condition would develop a causal model of the Irish population changes as they defend their position, with their essays including more connections of concepts and more causal connections than essays in the narrative and history conditions.

In addition to the essay task, students were asked to rate the importance of each statement of the presented text, with students rating the 70 statements either before or after writing the essay. This manipulation was carried out to test the hypothesis that students would write better essays if they first indicated the importance of the specific items of textual information.

It further hypothesized that the effect of rating the importance of items on essay quality would vary with the essay condition. In the argument condition, having students first indicate what is important would disrupt the development of the causal model by constraining the development of the model when it is being organized and written. Understanding should then be reduced in this condition. In the narrative and history essay conditions, however, such disruption would not occur because a causal model is not being developed. Moreover, it could be hypothesized that in the history and narrative essay conditions what students do may be to select the important contents and write about them, essentially the same thing that they may do when they are given an importance rating task. Finally, it was hypothesized that items students used in their essays would receive higher ratings than those not used, and by comparing the ratings given before and after essay

writing it would be possible to determine whether the writing of the essay produced changes.

Method

Ninety-six undergraduates at the University of Pittsburgh participated in this experiment for credit as part of an Introductory Psychology subject pool. All participants received information about Ireland from 1800 to 1850 in the form of eight separate sources, including a map; biographical accounts of King George III and Daniel O'Connell; brief descriptions of the Act of Union 1801, Act of Emancipation 1829, and the Great Famine; census data on the population size, the death rate, and the emigration rate between 1300 and 1850; and economic statistics on crop selling prices, rent costs, distribution of land holdings, and occupational breakdowns between 1800 and 1850.

An importance rating task was created by extracting the basic ideas from each of the sources (excluding the map) yielding 70 basic idea units. The idea units were listed on a page in random order. At the top of the page, students were asked "How important were the following points toward producing the significant changes in Ireland's population between 1846 and 1850?" They were then presented with a ten-point scale in which "1" was defined as "Not at all Important" and "10" as "Extremely Important." A short-answer 20-item general knowledge test was also given, the test containing questions such as "What did Gutenberg invent around 1450?" and "In what country did the Boxer Rebellion of 1900 occur?" Participants were given packets containing the separate sources about Ireland from 1800 to 1850. After reading through the information, one-half of the students in each essay condition were presented with the importance rating task and then a writing task. The other half performed the writing task before the importance rating task.

The writing task had the following instructions: "Historians work from sources including newspaper articles, autobiographies, and government documents like census reports to create histories. In this packet there are a number of documents about Ireland between 1800 and 1850. Your task is to take the role of historian and develop a *history* about what produced the significant changes in Ireland's population between 1846 and 1850. You will have about 30 minutes for this task. You are expected to make full use of that time."

One-third of the students saw the above instructions. For another third, the underlined word was replaced with *narrative* and for the remaining third the underlined word was replaced with *argument*. The resulting design is a 2×3 (task order $\times$ writing instruction) with 16 participants in each cell.

After completing both the importance rating and writing tasks, students completed a short questionnaire that requested information such as age, sex, educational status, number of college history courses taken, and amount of interest in history. They then completed a general history knowledge test. The students were in groups and each session lasted about one hour.

Results

Understanding of the presented material was assessed through analyses of the structure and content of the written accounts. As no differences were found in history knowledge across either the type of essay or the locus of the importance rating conditions, (Fs<1), this variable is not discussed further.

Analysis of students' written accounts

Three general aspects of student's writing were considered: the structure of their accounts, the integration of the information to be included in the accounts, and the selection of that information. Overall, analyses of the students' accounts indicated that both the essay type and timing of the importance ratings had an effect on the way students organized their written accounts and transformed and integrated the content within their accounts, especially with regard to causal relations. Specific analyses of students' writing included (a) the organization or structure of the account (i.e., collection of ideas versus causal essay), (b) the connection of idea units (i.e., the extent to which students recognized the possible relations between the ideas that were presented), (c) amount of explanation (i.e., number of causal connectives), (d) the origin of the information contained in sentences (i.e., information taken directly or paraphrased from sources, versus transformed or completely novel information), (e) the exhaustivity of the account (i.e., the extent to which the idea units mentioned in the text were included in students' accounts), and (f) importance of information included (i.e., the extent to which the idea units included in the essays were rated as the most important).

General description of the essays

The 30 minute writing task produced essays with an average of approximately 14 sentences. There were no differences in the length of essays due to writing task, $F = 1.19$, but there was a significant difference in length depending on whether the essays were written before or after the importance rating task. Students wrote longer essays when essays were written before the importance ratings were made, $M = 15.33$, than when written after the importance ratings were made, $M = 11.92$, $F(1,90) = 9.76$, $p < .002$. This result suggests that judging the importance of particular components of the passage acts to constrain the amount of information incorporated into the essay.

Measure of essay structure

Using Meyer's (Meyer, 1985) taxonomy, all essays were classified as either having a collective structure (that is, the essay consisted of a loosely or temporally associated listing of ideas) or a causal structure (that is, ideas of

Table 1 Observed frequencies of two text structures for writing task and task order conditions

		History	*Writing task narrative*	*Argument*
Task order				
Essay/Rating				
	Collective	7	7	1
	Causal	9	9	15
Rating/Essay				
	Collective	5	10	8
	Causal	11	6	8

essay were organized as antecedents and consequents). As shown in Table 1, both the type of essay and the timing of the importance rating procedure had an effect on the structures that were used. When essays were written before the importance rating task, almost all students in the argument writing condition wrote essays with a causal structure, while only about half the students in the narrative and history conditions did so. On the other hand, when students performed the importance rating task before writing their essays, no differences were seen across writing task conditions, with about half of the essays employing a causal structure. These results suggest that the importance rating procedure may not only have reduced the amount of information in the essay, but also likely had a disruptive influence on the construction of a causal model when it preceded the argument writing task.

Connections

The mean number of connectives used in writing an essay indicated that there were more connectives generated when essays were written before the importance ratings were made. $M = 10.27(4.86)$ than after they were made, $M = 8.15(4.06)$, $F(1,94) = 5.41$, $p < .02$. However, mean number of connectives was not a function of essay type, $F<1$. Also, while the interaction of the two variables is not significant, the p value is .13, $F(2,90) = 2.06$. When the essays were written before the importance ratings, the mean number of connectives written for argumentative essays was 11.00, for narrative essays was 10.94, and for history essays was 8.87. When the essays were written after the importance ratings were made, the means for the argumentative, narrative, and history essays were 6.75, 8.56, and 9.12 respectively. Thus, both argumentative and narrative essays contained more connections when the essay was written first. This was not true for the history essay condition.

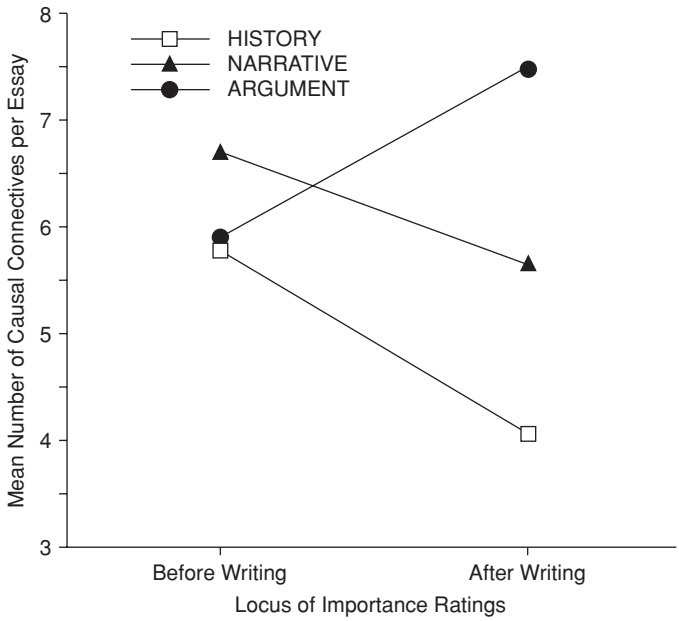


Figure 1 Mean number of causal connectives used in essays

Causal connections

For the number of causal connectives, whether the essay was written before or after the importance ratings did not matter, $F(1,94) < 1$. However, both essay type and the interaction of the two variables were significant. The number of stated causal connectives was greatest for the argumentative essay, $M = 6.69(3.47)$, with the narrative mean being 6.16(2.85) and the history mean 4.91(2.21), $F(1,94) =$, $p<.05$. A Tukey's test yielded a significant difference only between the argument and history conditions.

The significant interaction, $F(2,20) = 3.09$, $p<.05$, is shown in Figure 1. As indicated both the narrative and history conditions yielded more causal connectives when the essay was written after the importance rating task, whereas the argumentative essay condition yielded more causal connectives when the essay was written first. These results suggest that the argumentative essay writers were disrupted and constrained in making causal links when writing an essay if they were required to first state importance ratings. While the previous result suggests that individuals writing argumentative essays develop causal models, this interaction suggests that performing the ratings emphasizes the text contents as isolated bits of information which acts to prevent integration. Moreover, since the presented material was available while the essays were being written, the reduction in the number of causal connectives cannot be seen as a matter of memory.

As noted for the narrative and history conditions, however, the essays contained more causal connectives when the importance ratings were made before the essay was written. This result indicates that in these conditions the rating procedure facilitated the use of causal connectives, suggesting that perhaps the strategy used in these conditions was to determine what is important and then to fill in the gaps between the important events or actions. The data suggest that this tendency occurs more in the narrative than in the history condition.

Transformations

Extending a distinction made by Greene (1994), the essays were analyzed in relation to the origin of the contents. Coded on a sentence basis, a sentence is termed "borrowed" if the contents are taken directly from the presented text, "added" if the contents were not in the presented text, and "transformed" if the sentence contains a combination of "borrowed" and "added" information or if it combines information presented in the text in a new way. The proportion of transformed sentences was .50 for the argumentative essay group, .33 for the narrative group and .40 for the history group, $F(2,90) = 7.27$, $p < .001$. A Tukey's test indicated that the argument essays contained a significantly greater proportion of transformed sentences than the narrative essays.

Exhaustivity

Exhaustivity was measured by the number of idea units out of the 70 in the presented material that were included in the essay. More items were used in essays when the essays were written first, before the importance ratings $M = 17.71(6.08)$, than when the essays were written after the importance ratings, $M = 14.94(5.21)$, $F(1,94) = 6.01$, $p < .02$. Essay type was not significant, $F(2,93) = 1.05$, while the interaction of the two variables was significant at the .09 level, $F(2,90) = 2.43$. This near significance is reflected primarily in the argumentative essay condition, in which the number of idea units included when the essay was written before the importance ratings was 19.62, but when the essay was written after, the mean was 13.31. The corresponding means for the narrative and history conditions were 18.19 and 16.50 (before) and 15.50 and 15.00 (after). This result is consistent with the idea that the rating procedure tended to inhibit the development of argumentative essays.

Importance ratings

The mean overall importance ratings did not vary with respect to when importance ratings were made, $F < 1$, essay type, $F < 1.28$, or the interaction

of the two variables, $F<1$. However, those items of the text that were used by a given writer in an essay were rated as more important than those not used in the essay $F(1,90) = 165.31$, $p<.0001$, with the mean of used items 7.76 and unused items 5.21. While there was little difference in importance ratings of unused items related to when the ratings were made, used items were rated as significantly more important when ratings were done after essays were written, $M = 8.15$, than before, $M = 7.37$, $F(1,90) = 4.73$, $p<.03$. This result indicates that using items in an essay increased the perceived importance of those items.

While essay type and the interaction between essay type and the timing of essay were not significant, Fs<1, there was a significant interaction between whether an item was used or not used and essay type, $F(2,90) = 5.26$, $p<.007$. The means for the three essay types for the not-used items were narrative 5.45, argumentative 5.29 and history 4.91, while for the used items the respective means were 7.13, 8.03, and 8.13. Thus, the narrative essay condition had the highest rating for unused items and the lowest rating for the used items, suggesting that individuals in the narrative condition tended to include more less-important statements in the writing of their essays, since essay type did not yield significant differences in the overall importance ratings. The history condition however had the lowest unused ratings and the highest used rating, suggesting that individuals writing the history essay were especially concerned with including the most important information in their essays.

The importance rating data, therefore, supported the hypotheses that in writing essays individuals tended to use the items they considered to be most important, and when an item is used it is perceived as more important than it is before it is used.

Discussion

The present study replicated some of the findings of an earlier study (Wiley & Voss, 1996), namely, that individuals who write arguments from sources (without having made importance ratings) produced essays that included more causal connections and transformations of presented material than individuals who wrote a narrative or a history. Essays written in the argument condition were also more likely to have a causal instead of a collective, list-like structure than essays written in the narrative and history conditions. These results, again, suggest that individuals in the argument condition are constructing causal models to a greater extent than the individuals in the other two conditions.

Further, the present study demonstrates that an importance rating task can disrupt the building of causal models in the argument condition. When the importance rating task preceded the essay task, the argument condition no longer had significantly more connections, causal connections,

or transformations than the other two conditions. Nor were there more causal structures than collective ones in the argument condition when the rating task preceded essay writing. Thus, rating the importance of specific items may constrain the integration and organization of text information, and the relating of it to information in memory. It also reduced the number of concepts employed overall in essay writing.

An interesting twist is that although the importance rating task seemed to disrupt the integration of information in the argument condition, no decreases were seen in the other two conditions. In fact, there is some evidence that the importance rating task may have facilitated the integration of source information in the narrative and history conditions, as the number of casual connectives was greater in these conditions when importance ratings were made first than when they were after writing the essays.

The present results relate to the distinction that has been made between "knowledge-telling" and "knowledge-transforming" writing processes (Bereiter & Scardamalia, 1987) or more generally to the amount of constructive activity that occurs during writing. Bereiter and Scardamalia have described the writing process of more and less skilled writers, finding that while less skilled writers often rely on the explicit assigned topic or genre to structure their writing, more skilled writers are more likely to have a main point or structure "emerge" from their thinking about the textual information (Bereiter, Burtis, & Scardamalia, 1988). And, whereas less-skilled writers use textual information to fill in the "slots" of a generic discourse schema, more skilled writers create novel frameworks based on consideration of both the specific content and task, and then adapt or "transform" the textual information to cohere within those frameworks. Hence, the less skilled writers are basically reporting what they have been told, perhaps with some surface re-arrangement, in a process involving little constructive activity that has been seen as "knowledge-telling." On the other hand, more skilled writers select, re-organize, integrate, and synthesize textual information, in a process involving a great deal of constructive activity that has been called "knowledge-transforming."

One way of viewing the present study is as an attempt to find conditions under which students are more likely to engage in "knowledge-transforming" or constructive activity while writing. Our results suggest that the argumentative essay task provided for more knowledge transformation than the other two essay instructions. This seems to be because developing an argument involves selecting and organizing information as well as transforming it and relating it to information in memory. This process can be quite a constructive and creative act. While writing a narrative also may be, students in the Wiley and Voss (1996) study writing from a text already had the narrative present, while those writing from sources in both studies apparently emphasized the chronological component of a narrative rather than the causal

structure it can afford. Hence, the narrative task in this study led to more "knowledge-telling."

The importance rating results can also be viewed from a similar perspective. It may be that the importance rating task emphasized the processing of text information on an item-by-item basis, which is often an attribute of low constructive activity (Chan, Burtis, Scardamalia, & Bereiter, 1992). This would be consistent with our results that the importance rating task lessened the amount of "transformation" that was seen in the argument writing condition, but may have helped students who were not already attending the information on a more global level (as in the narrative and history conditions).

The present results support the general theoretical orientation taken in this chapter that the understanding of history can be facilitated by certain learning contexts, specifically when students write arguments from a number of sources. Further, the results suggest that an importance rating task may improve understanding in some conditions where little understanding would be developed otherwise, but importance rating may hinder the development of a causal model in conditions where it would be well developed. Furthermore, one could generalize these findings to the possible disruptive influence that an emphasis on facts may produce in the development of understanding. (See also Leinhardt, Stainton, Virji, & Odoroff, 1994 and Voss & Wiley, in press.)

These results, taken together with the results of the previous experiment (Wiley & Voss, 1996), offer insights into contexts that may yield better understanding in the history classroom by prompting the integration and transformation of presented information. A further advantage to a learning context that involves writing an argument is that it is more likely to produce text "ownership" for the writer. Students often see history as "someone else's facts" (Holt, 1990). By writing their own arguments, students may begin to see that history is not just about learning names and dates, but an on-going debate about what those facts may mean. In addition to providing a more motivating context, instruction that emphasizes the historian's role of integrating and interpreting isolated facts into explanatory arguments has the potential to improve not only understanding of specific historical content, but the understanding of history as a subject matter as well.

Acknowledgements

The authors thank Laurie Silfies for her research assistance. The research reported in this paper was supported by the Office of Educational Research and Improvement of the United States Department of Education via an award of the Center for Student Learning to the University of Pittsburgh. The views expressed in this paper are not necessarily those of any of these organizations.

References

Beatty, A. S., Reese, C. M., Persky, H. R., & Carr, P. (1996). *NAEP 1994 U.S. history report card*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Bereiter, C., Burtis, P. J., & Scardamalia, M. (1988). Cognitive operations in constructing main points in written composition. *Journal of Memory and Language*, *27*(3), 261–278.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Carretero, M., & Voss, J. F. (Eds.) (1994). *Cognitive and instructional processes in history and the social sciences*. Hillsdale, NJ: Erlbaum.

Chan, C. K. K., Burtis, P. J., Scardamalia, M., & Bereiter, C. (1992). Constructive activity in learning from text. *American Educational Research Journal*, *29*(1), 97–118.

Greene, S. (1994). Students as authors in the study of history. In G. Leinhardt, I. L. Beck, & C. Stainton (Eds.), *Teaching and learning in history* (pp. 137–170). Hillsdale, NJ: Erlbaum.

Holt, T. (1990). *Thinking historically: Narrative, imagination, and understanding*. College Entrance Examination Board: New York.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294–303.

Leinhardt, G., Stainton, C., Virji, S. M., & Odoroff, E. (1994). Learning to reason in history: Mindlessness to mindfulness. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 131–158). Hillsdale, NJ: Erlbaum.

Meyer, B. J. F. (1985). Prose analysis: Purposes, procedure, and problems. In B. Britton & J. Black (Eds.), *Understanding expository text* (pp. 11–64). Hillsdale, NJ: Erlbaum.

Ravitch, D., & Finn, C. (1987). *What do our 17-year-olds know?* New York: Harper & Row.

Voss, J. F., & Wiley, J. (in press). Conceptual understanding in history. *International Journal of Educational Research*.

Voss, J. F., Wiley, J., & Carretero, M. (1995). Acquiring intellectual skills. *Annual Review of Psychology*, *46*, 155–181.

Wiley, J., & Voss, J. F. (1996). The effects of "playing historian" on learning in history. *Applied Cognitive Psychology*, *10*, 563–572.

78

GENERATIVE TEACHING

An enhancement strategy for the learning of economics in cooperative groups

M. Kourilsky and M. C. Wittrock

The purpose of this study was to increase the learning of economics among lower socioeconomic level public high school students by teaching them to use generative comprehension procedures in their economics classes' cooperative learning groups. In a randomly assigned two-treatment design, it was predicted and found that generative learning procedures in cooperative learning classes increased ($p < .0001$) the learning of economics by sizable amounts compared with a control procedure that used only cooperative learning methods and that produced smaller increases. Students' confidence in the correctness of their answers increased ($p < .0001$), and the level of misinformation decreased ($p < .0001$) as a result of generative teaching procedures. These facilitative effects of generative teaching occurred for both males and females.

The economic illiteracy of our nation's youth—especially our at-risk youth—forges a compelling national consensus for action that has been reflected in 28 state mandates for economics education before high school graduation. In addition to the widely documented shortage of teachers with the minimum requisite knowledge of economics, teachers with satisfactory knowledge of economics succeed in imparting only rudimentary levels of learning and often fail to teach economics at the higher levels of cognition: comprehension, analysis, and synthesis. As late as 1988, when 26 of the state mandates regarding the teaching of economics were in effect and teacher training of prospective high school teachers of economics was accelerating rapidly, the testing of 8,000 randomly selected high school juniors and seniors throughout the nation revealed a low level of economic literacy. The average score was 40% correct on a test that consisted predominantly of basic and simple economics concepts (Walstad & Soper, 1988).

Source: *American Educational Research Journal*, 1992, 29(4), 861–876.

High school graduates should be able to think, critically analyze, and assess. As voters, they will make decisions that will direct the agenda of our nation.

> Economics stands out among disciplines because it teaches theory, requires analysis, and blends the rigors of science with the necessities of human welfare. Economics grapples with the realities of change and requires the learner to adapt to new and dynamic phenomena. . . . An understanding of economics will provide the decision-making capabilities that will enable the United States to respond to this dynamically changing world . . . and to further our progress into the future (Brenneke, 1992).

As Nobelist Paul Samuelson eloquently summarizes, "All your life—from cradle to grave and beyond—you will run up against the brute truths of economics" (Samuelson, 1983 p. ii).

The primary purpose of this study was to increase the learning of economics among lower socioeconomic level public high school students by teaching them to use generative comprehension procedures in their economics classes' cooperative learning groups. Cooperative learning provides a group learning goal, the successful achievement of which often hinges on the stimulation of student interest in each other's learning processes. The comprehension procedures of generative teaching, which emphasize the importance of student construction and revision of preconceptions in the acquisition of concepts (Wittrock, 1974, 1990, 1991), complement the motivational advantages of cooperative learning. Generative teaching gives the motivated students in cooperative learning groups effective ways to discover and, when necessary, revise the preconceptions about economics of their colearners.

Cooperative learning and generative teaching have individually been shown to be effective for enhancing classroom learning. In Slavin's synthesis (1990), he reports that by coupling a common group goal with individual accountability for learning, cooperative learning enhances academic achievement in school.

Over the past few years, cognitive psychologists have amassed considerable evidence that learning is an active process. The student is not a passive recipient of the teacher's instruction (Gagné, 1985; Gall, Gall, Jacobson, & Bullock, 1990). Rather than linking teacher behavior directly to student achievement, a cognitive perspective maintains that teaching influences student thinking, which in turn mediates learning and achievement (Wittrock, 1986). In other words, students' cognitive and affective thought processes contribute to their ability to comprehend and master academic subject-matter content.

The theory of generative learning explains the process with which students transform the unfamiliar into the familiar by generating their own

connections from that which is already understood to that which is to be learned. Generative teaching involves knowing the learners' preconceptions of the subject matter and leading the learners to revise these preconceptions by teaching them to generate meaning from instruction. Generative teaching also seeks to foster a distinctive type and quality of student motivation (and attention) that emphasizes taking control and responsibility for being active and attentive in learning. Examples include encouraging students to actively engage in classroom learning activities and to attribute their success and failures to their own generative efforts.

Generative teaching facilitates student learning by (a) attending to students' knowledge, preconceptions, attention, motivation, acquisition, and comprehension; and (b) directing students to construct relations among the subject-matter concepts to be learned and between these concepts and their knowledge or experience. These generations enable the students to revise their preconceptions and to construct meaning and understanding from instruction.

Generative teaching regularly shows positive results with respect to the learning of subjects taught in schools, including economics, reading, mathematics, science, and geography. Kourilsky and Wittrock (1987) found that comprehension of economics in senior high school classes was facilitated by a generative teaching strategy that initially presented the concepts in a familiar verbal mode and then presented them in a less familiar spatial mode using graphs. Also, in the context of a mentor teacher training institute, Kourilsky (in press) discovered that the generative teaching strategy of identifying and correcting "incorrect mind-sets" of teachers resulted in large reductions in the teachers' economic misconceptions. Wittrock, Marks, and Doctorow (1975) doubled reading comprehension and retention by giving fifth and sixth graders familiar contexts to use to generate meaning. Reading comprehension of elementary school children doubled by use of a generative procedure of teaching children verbal and spatial strategies for relating the text they read to their everyday experiences (Linden & Wittrock, 1981). At the junior high school level, students improved their reading comprehension by about 100% by generating their own summaries for each paragraph of a story text they read (Doctorow, Wittrock, & Marks, 1975). At the college level, the use of generative teaching procedures resulted in sizable increases in reading retention and comprehension (Wittrock & Alesandrini, 1990). Generative learning has also been studied in mathematics (Peled & Wittrock, 1990), science education (Osborne & Wittrock, 1983, 1985), and geography (MacKenzie & White, 1981).

The authors believed that by combining generative teaching and cooperative learning, we would enhance the contributions of each of them to student learning. When a teacher employs cooperative learning as an instructional strategy, he or she tacitly asks students in the group to act as teachers for their group peers. Cooperative learning stresses the social context of learning, whereas generative teaching addresses the cognitive processes of constructing

meaning by relating experiences and knowledge to instruction. Training students in generative teaching provides an effective subject-matter-oriented educational methodology to individuals who will subsequently be engaged in peer teaching. On the other hand, for an instructor to be effective in generative teaching, he or she must know and address the preconceptions of the learners. If these preconceptions are actually misconceptions, the teacher must lead the learners to change the incorrect mind-sets that led to false preconceptions. Each student in the cooperative learning groups, by virtue of peer linkages and shared experience, can learn how the others think and may be more likely than some classroom teachers to discover and modify peers' preconceptions. Finally, in the process of exploring their peers' preconceptions and misconceptions, the students in the cooperative groups can develop a heightened sensitivity to their own preconceptions. Therefore, the authors maintain that by integrating generative teaching strategies and cooperative learning, an increase in the understanding of economics and a reduction in the misconceptions about the subject matter will result, compared with cooperative learning procedures alone. In addition, the authors believe that the generation of links between the instruction and the learners' mind-sets increases the learners' confidence in their comprehension of the concepts of economics.

In this study we explored the effects of generative teaching upon the comprehension of economics by middle and lower socioeconomic 12th-grade students enrolled in economics classes taught by cooperative learning methods. To provide a careful test of the model in a realistic teaching setting, each of the students in this study was individually assigned at random either to the generative teaching treatment or to the control treatment, which differed from each other only in the following way: In the generative teaching condition the students learned how to use generative procedures to identify and to revise one another's conceptions of economics; however, in the control condition, the students were given the same amount of time to learn and to review principles of cooperative learning. This review was deliberately redundant with the cooperative learning procedures all students in this study had been using daily in their economics classes. The students' regular economics teacher presented all economics instruction in both treatments of this study. Learning time was held constant across treatments at 17 hours. Within that experimental design, we tested the following hypotheses about generative teaching:

1. Learners instructed in the Model of Generative Teaching *learn more economics* (both graphical and verbal understanding) than do learners not instructed in the Model of Generative Teaching.
2. Learners instructed in the Model of Generative Teaching will be *less misinformed in economics* than those who do not receive instruction in the Generative Model of Teaching.

3. Learners instructed in the Model of Generative Teaching will have *more confidence in correct economics information* than learners not instructed in the Model of Generative Teaching.

We were also interested in ascertaining whether gender differences occurred with respect to economics understanding, confidence in correct information, or percentage of misinformed responses. In previous research, Kourilsky and her colleagues (Kourilsky & Campbell, 1984; Kourilsky & Oritz, 1985; Kourilsky & Graff, 1986) found no gender differences either in verbal comprehension or in graphical comprehension of economics among elementary school children. Although we anticipated that in general there would be no gender differences among secondary students, there is some research that suggests that gender differences on spatial/graphical tasks sometimes develop during adolescence, in favor of males (Cochran & Wheatley, 1989); Fralley & Eliot, 1976; MacCoby & Jacklin, 1974). For this reason we decided to include gender as a variable to continue to gather data on male-female differences during adolescence. We tested the following hypothesis:

4. There will be no gender differences in graphical or verbal economic understanding, confidence in correct information, or level of misinformation in either the experimental group or the control group.

Method

Participants

One hundred and forty-two high school seniors (12th graders) from a lower-to-middle socioeconomic level neighborhood in El Segundo, California, comprised the sample. There were 73 males and 69 females.

Tests

The Market Equilibrium Test (Forms A and B), a 45-minute test, consists of 20 items on the third unit of a five-unit semester course. This unit concerns functions of a market system and the rationing function of price. The test includes items on demand, supply, market price, price ceilings, and price floors. There are 7 verbal and 13 graphical items. All questions on this test are at or above the comprehension level of cognition (Bloom, 1956).

The Market Equilibrium Test was constructed by a team of four subject-matter experts and independently reviewed by a panel of 10 economic educators for content validity and test-item equivalence between Form A and Form B. The panel found that each of the test items was a valid measure of its objective. Additionally, the tests were matched item-per-item. The panel agreed that each item on Form A was equivalent to its counterpart on Form

1. "Economic demand" for a product refers to:
   (a) how much of the product the people are willing and able to buy at alternative prices?
   (b) how many people are willing and able to buy the product at alternative prices?
   (c) how much of the product people want, whether they can afford to buy it or not?
2. The graph at the right presents a supply curve of chocolate. Which of the following explains the shift of the supply away from Old Supply to New Supply?
   (a) a higher price of chocolate;
   (b) a decrease in consumers' incomes;
   (c) a higher price of cocoa, which is used to produce chocolate.

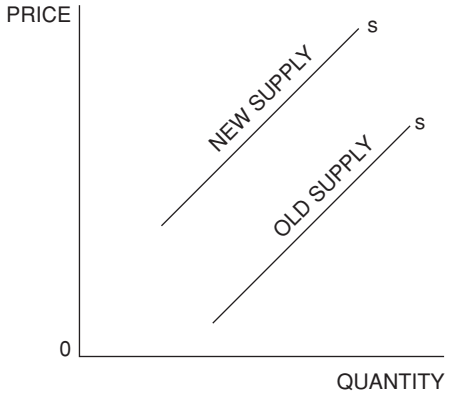
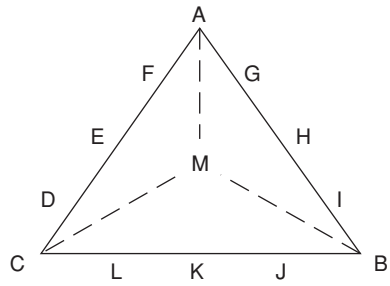


Figure 1 Sample test items

B in terms of difficulty and the content being measured. (See Figure 1 for sample questions and Figure 2 for the scoring procedure.) To ascertain test-retest reliability, the Market Equilibrium Test was given to 80 12th graders in a neighboring school district. Forty randomly chosen students took Form A, and another group of 40 randomly chosen students took Form B. A week later the same groups repeated the test on Form A and Form B, respectively, to determine test-retest reliability. The result was a test-retest reliability of $r = .93$ (Form A) and .95 (Form B).

Additionally, as a cross-check of Form A and Form B equivalence, the differences in mean scores between the Week 2 group on Form A and the Week 2 group on Form B were calculated and found to be statistically nonsignificant.

The design of the Market Equilibrium Test accommodates the Information Reference Testing (IRT) scoring procedure (Bruno, 1986). The IRT scoring procedure evaluates students on their information and on their confidence in their information. The scoring of the students' tests incorporates this confidence weighting to obtain a more reliable representation of the information state of the student, i.e., fully informed, partially informed, uninformed ("I dont't know"), or misinformed ("complete confidence in incorrect information"). The IRT eliminates the bias generated from guessing,



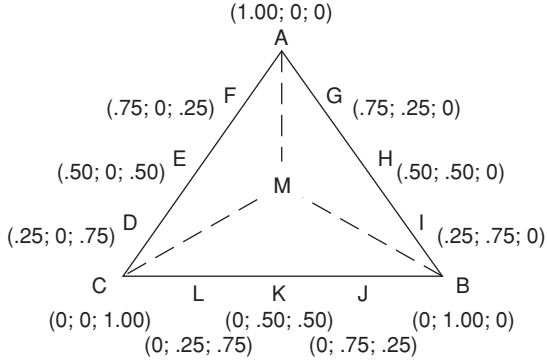
The IRT Response Triangle

The scaling factors A (–63.12) and B (33.00) in the IRT log formula generate the following awards for confidence in the correct answer.

Confidence	Actual	Approximate Score for Use in the Classroom	Interpretation of Information State
1.00	30.12	+30	Informed
.75	22.23	+20	Near Informed
.50	11.12	+10	Part Informed
.33	–.27	0	Uninformed
.25	–7.88	–10	Near Misinformed
0.00	–99.01	–100	Misinformed

IRT Point Awards

Conditional score triplets can then be associated with each response option on the IRT triangle.



Conditional Probability Triplets on the IRT Response Triangle

Figure 2 IRT response triangle

and allows for partial credit when a student correctly narrows his or her choice of possible correct answers. With the IRT format, point awards are determined by using the formula: Score = $A \log P_i + B$ where A and B are scaling factors and P_i is the student confidence in the correct answer. The

scaling factors are needed to insure that a score of 0 is awarded a P_i of .33 and to promote ease of communication (and remembering) to students. Figure 2 presents a sample of the IRT response schedule and scoring procedure.

The student selects among response options A-M, depending upon both his or her choice of the correct response and confidence in that choice. For example, if a student chose letter A on question 1 (see Figure 1), it would indicate that he or she is *fully informed* (100% sure and correct) and confident in the correct response. If a student on question 1 chose letter C, it would indicate that he or she is *misinformed*, completely confident in incorrect information (100% sure and incorrect). If a student chose letter H (in-between A and B), it would suggest that he or she believed that C is incorrect, and the correct answer is either A or 3 (50% sure). Such a student would be *partially informed*. A student who is *nearly sure* of A (75% confidence) would respond with letters F or G. If a student is *uninformed* (33% confidence in each choice), he or she would respond with letter M. The IRT new scores range from –100 to +30 on *each* test item.

Procedures

All the economics students in the 12th grade were individually assigned at random to six classes in economics. Each class was then randomly as signed either to the experimental group or to the control group. Students within the experimental and control groups were then individually randomly assigned to cooperative learning groups consisting of four members each. To determine the equivalence of the experimental and control groups, each student took Form A of the Market Equilibrium Test as a pretest. Each student received the results of his or her pretest, including the total score and a breakdown of each concept on which he or she was fully informed, partially informed, uninformed, or misinformed. No statistically significant differences in the pretest scores occurred across gender or across experimental and control groups. All participants in the study had previous experience and instruction on both IRT testing procedures and cooperative learning.

The experimental group received 2 hours of training on generative learning by an expert in teacher education, whereas the control group received 2 additional hours of review on cooperative learning strategies that had been presented to both the experimental and control groups. *The following is a summary of the Generative Teaching Treatment*:

Step 1. Students discussed and were given a definition of Generative Teaching Principles, particularly the importance of attending to four factors: (a) one another's preconceptions, knowledge, and perception; (b) motivation; (c) attention, and (d) generation. The students were also given a list that included the following: (a) relate the subject matter presented in class to the learners' prior knowledge (preconception); (b) relate subject matter to the learners' beliefs, preconceptions, and real world experiences (preconceptions);

(c) use visual as well as verbal examples (generation); (d) take one another's learning styles into account (preconceptions, generation); (e) ask questions that direct each other's attention to the major content to be learned (attention); (f) ask each other high level questions, not recall questions (attention); (g) have high expectation levels for everyone in the group (motivation); (h) have everyone in the group periodically make summaries in their own words (generation); (i) make sure everyone in the group is actively involved in the learning (motivation, generation); (j) remind each other that each person is ultimately responsible for his or her own learning (motivation).

Step 2. Students then were given a definition and explanation of each of the three types of incorrect mind-sets in learning economies that have been shown to result in economic misconceptions: (1) a *linguistic mind-set*, which derives from natural language usage and the subsequent psychological tendency to identify with the natural language use of the term or concept; (2) a *physical mind-set*, which derives from the individual's physical experience, which then leads to an incorrect physical analogy; and (3) a *resistive mind-set*, which derives from the natural resistance to acknowledge a reality that is in conflict with what the individual believes *ought to be* and the subsequent tendency to psychologically ignore or deny that reality.

Step 3. Students then were given examples of each of these incorrect mind-sets in terms of the economics concepts they had previously learned. For example, in learning scarcity, some of the students had trouble comprehending that scarcity is a relative concept and instead used it as a synonym for rarity; some students in learning opportunity cost believed it was the sum of all alternatives foregone as opposed to the second-best alternative.

Step 4. Students then were asked to generate their own examples of incorrect mind-sets in previously learned economics concepts and relate how they succeeded in replacing these incorrect mind-sets with correct mind-sets.

Step 5. Students then were challenged in the context of their cooperative learning groups: (1) to "think out-loud" with respect to the concepts in the current learning sequence on which they were misinformed and to actively "get into each other's minds"; (2) to identify in one another incorrect mindsets that were preventing total comprehension (e.g., thinking of *demand as adamantly desired* or *insisted upon* as opposed to *desire backed by the ability to pay*); to help one another correct the incorrect mind-sets by incorporating the principles of motivation, attention, and generation.

Each group then received 15 hours of instruction in economics using cooperative learning methods. The same high school economics teacher presented the instruction for the experimental and control groups. However, most of the dialogue, about 80%, was conducted by the students themselves in their cooperative learning groups.

Two economic educators trained in both generative teaching strategies and cooperative learning methods were present throughout the instruction in both treatments. They evaluated compliance with the Model of Generative

Teaching and reported that the individuals in the experimental treatment regularly (about 75% of the time) invoked generative teaching strategies in their cooperative learning groups.

At the end of the 3-week instructional period, all participants were tested on their verbal and graphical understanding of economics, their levels of misinformation, and their levels of confidence in correct information. Both experimental and control groups took Form B of the Market Equilibrium Test. Again, 45 minutes were allotted for this test.

The study was "double blind" in the sense that neither the teacher nor the students were aware of the hypotheses or design of the experiment.

Results

To test the effectiveness of the Model of Generative Teaching on graphical and verbal economics understanding and to estimate the gender effect, a two-way factorial analysis of covariance (gender by treatment group) was used for each of the two dependent variables—graphical understanding and verbal understanding. The graphical understanding score included all items that required understanding a displayed graph to answer the question. The verbal understanding score included items that did not require understanding of a displayed graph to answer the question.

The covariates were the pretest scores on graphical and verbal understanding. Analysis of covariance (ANCOVA) was used to test the significance of the effect of the experimental treatment on the posttest scores on graphical and verbal understanding adjusted for the pretest performance. As indicated in Table 1, in the case of graphical understanding, the treatment effect was significant [F (1, 137) = 21.75, $p < .0001$]. The gender effect and the interaction effect between gender and group were not significant. An R-square of .38 was obtained, indicating that 38% of the variance in graphical understanding was accounted for by the treatment.

Table 2 indicates for the verbal understanding the treatment effect also was significant [F (1, 137) = 13.81, $p < .0003$]. Neither the gender nor the interaction effects were significant. The R-square was equal to .28.

To examine the effects of generative teaching on the subjects' level of misinformation, a two-way factorial analysis of covariance (gender by treatment) was used with level of misinformation as the dependent variable. The covariate was the pretest score on the level of misinformation. ANCOVA was used to test the significance of the treatment on the posttest scores of males and females. As shown in Table 3, a statistically significant treatment effect was found [F (1, 137) = 20.88, $p < .0001$]. No significant gender or interaction effects were revealed. The R-square was .34.

The same type of factorial analysis of covariance was employed to ascertain the effect of treatment and gender on the participants' confidence in correct information. Again there was a significant treatment effect [F (1, 137) = 35.21,

Table 1 Pretest and posttest mean scores (and percentages of possible gain) for graphical economic understanding by gender and treatment

Group	*n*	*M*	*SD*	*LS M**	*Percentage of possible gain*
			Experimental		
Male	43				
Pretest		−219.53	272.47		
Posttest		90.23	295.84	101.28	51
Female	33				
Pretest		−184.85	258.19		
Posttest		70.91	233.61	62.82	45
			Control		
Male	30				
Pretest		−234.67	302.46		
Posttest		−105.67	275.11	−86.27	21
Female	36				
Pretest		−159.72	280.52		
Posttest		−87.78	280.29	−109.73	13

* LS = least square

Table 2 Pretest and posttest mean scores (and percentages of possible gain) for verbal economic understanding by gender and treatment

Group	*n*	*M*	*SD*	*LS M*	*Percentage of possible gain*
			Experimental		
Male	43				
Pretest		−183.95	187.77		
Posttest		16.74	186.89	14.31	51
Female	33				
Pretest		−210.91	157.73		
Posttest		−42.42	185.05	−31.92	40
			Control		
Male	30				
Pretest		−167.33	167.35		
Posttest		−140.43	154.97	−150.73	7
Female	36				
Pretest		−190.51	178.03		
Posttest		−73.06	193.41	−71.11	29

Table 3 Pretest and posttest scores for level (and percentages of possible reduction) of misinformation in economics by gender and treatment

Group	*n*	*M*	*SD*	*LS M*	*Percentage of possible gain*
		Experimental			
Male	43				
Pretest		646.51	345.96		
Posttest		272.09	293.04	262.30	58
Female	33				
Pretest		600	379.97		
Posttest		309.09	260.25	317.74	48
		Control			
Male	30				
Pretest		656.67	403.16		
Posttest		523.33	284.88	509.52	20
Female	36				
Pretest		583.33	390.24		
Posttest		452.78	338.47	468.04	22

Table 4 Pretest and posttest mean scores (and percentages of possible gain) for confidence in economic understanding by gender and treatment

Group	*n*	*M*	*SD*	*LS M*	*Percentage of possible gain*
		Experimental			
Male	43				
Pretest		.248	.127		
Posttest		.571	.316	.58	43
Female	33				
Pretest		.237	.075		
Posttest		.474	.272	.49	31
		Control			
Male	30				
Pretest		.276	.174		
Posttest		.324	.208	.30	7
Female	36				
Pretest		.260	.105		
Posttest		.359	.217	.35	13

$p < .0001$], and no significant gender or interaction effects. An R-square of .44 was obtained. (See Table 4.)

The statistical analyses of the data support the four hypotheses. Subjects instructed in the Model of Generative Teaching comprehended more economics ($p < .0001$), were less misinformed with respect to economic understanding ($p < .0001$), and had more confidence in correct information ($p < .0001$) than did those subjects who did not receive instruction in the Model of Generative Teaching. There were no gender differences in economic understanding (graphical or verbal), misinformation, or confidence in correct information.

Discussion

This experiment explored three predictions of the Model of Generative Teaching about the learning and comprehension of economics concepts by 12th graders in economics classes taught by generative teaching procedures in a cooperative learning context. The Generative Treatment differed from the Control Treatment in the ways that students learned to construct meaning for concepts in economics. In the Generative Treatment, the students learned to generate relations (a) across the concepts of economics and (b) between these concepts in economics and their experiences.

As reported by the two economic educators who monitored the administration of each treatment, the students in the Generative Treatment learned to identify one another's misconceptions and to generate alternative relations between the concepts in the text and their experiences, not between these concepts and the students' misconceptions. Cooperative learning appeared to provide an excellent context for student discovery of one another's misconceptions and for the generation of alternative relations that better synthesize experience within the concepts of economics.

The first hypothesis stated that generative teaching in a cooperative learning classroom context increases the comprehension of economics, compared with a procedure that employs only cooperative learning. The first hypothesis was supported ($p < .0001$). The experimental group significantly outperformed the control group through instruction that occurred in the students' regular classrooms with the students' regular economics teacher. Under these realistic conditions, and with no additional amount of time to learn, generative teaching showed a highly statistically significant gain, the magnitude of which bodes well for its real-world meaning and practicality. The large gain in learning involved minimal added cost and only 2 hours of instruction in generative learning procedures.

The second hypothesis predicted that generative teaching reduces the students' misinformation or erroneous conceptions of the principles of economics taught in the experiment. The data again clearly supported the hypothesis ($p < .0001$). The level of misinformation decreased significantly more in the

experimental group than in the control group. Apparently the instruction in learning to recognize and to modify each other's preconceptions about principles of economics influenced the revision of some of these misconceptions into more useful or more sophisticated conceptions of economics. These relearnings of concepts of economics appear to be involved in the increase in economics comprehension that generative teaching produced in this study.

The third hypothesis predicted that generative teaching increases students' confidence in their correct answers. The data also support this hypothesis ($p < .0001$). Confidence in correct responses was significantly higher for the experimental group than for the control group. In the experimental group, this large gain in confidence in correct responses accompanied the large increase in economics comprehension and the sizable decrease in misinformation about the economics concepts taught in class. These findings also imply that generative teaching enhances students' awareness of and confidence in their increase in economics learning. If that is the case, then generative teaching may have positive effects upon student attributions for learning and upon their conceptions of their ability to learn economics. In addition, generative teaching may improve the students' willingness to rely on their economics understanding to analyze critically and perhaps challenge rhetorically persuasive economics agendas that are not based on valid economic assumptions or reasoning. These possible effects deserve study in future research.

The fact that there were no gender differences in economic understanding, misinformation, or level of confidence in correct information suggests that the generative treatment is nongender preferential and is capable of achieving significant results across both genders.

In sum, the results of this experiment clearly supported all the predictions of the Model of Generative Teaching. Because we conducted these experiments in a public school, in which 15 hours of instruction in economics occurred in the students' regular classes that were taught by the students' regular economics teacher, the results of this study gain some external validity with respect to enhancing the teaching of economics in a realistic secondary school setting.

In essence, by training the students in generative teaching strategies in this experiment, we increased their effectiveness as learners and their effectiveness as teachers of their peers in the cooperative learning classes. The sizable increase in learning obtained by generative strategies necessitated no special economics curriculum and no added time to learn economics. The increase in learning involved teaching students effective ways to revise their own and their peers' understanding of economics by generating meaningful relations between their knowledge, beliefs, and experiences and the principles taught in their economics classes.

References

Bloom, B.S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Company.

Brenneke, J.S. (1992). The case for economics education. In J.S. Brenneke & F.W. Rushing (Eds.), *An economy at risk* (p. 3). Atlanta: Georgia State University Business Press.

Bruno, J.E. (1986). Assessing the knowledge base of students: An information theoretic approach to testing. *Measurement and Evaluation in Counseling and Development*, *19* (3), pp. 116–130.

Cochran, K.F., & Wheattley, G.H. (1989). Ability and sex-related differences in cognitive strategies on spatial tasks. *Journal of General Psychology*, *116* (1), 43–55.

Doctorow, M.J., Wittrock, M.C., & Marks, C.B. (1975). Generative processes in reading comprehension. *Journal of Educational Psychology*, *70*, 109–118.

Fralley, S.S., & Eliot, J. (1976). Sex differences in spatial abilities. *Young Children*, *31* (6), 487–497.

Gagné, E.D. (1985). *The cognitive psychology of school learning*. Boston: Little, Brown, & Company.

Gall, M.D., Gall, J.P., Jacobsen, D.R., & Bullock, T.L. (1990). *Tools for learning: A guide to teaching study skills*. Alexandria, VA: Association for Supervision and Curriculum Development.

Kourilsky, M.L. (in press). Economic education and a generative model of mislearning and recovery. *Journal of Economic Education*.

Kourilsky, M.L., & Campbell, M. (1984). Sex differences in a simulated classroom economy. *Sex Roles: A Journal of Research*, *10*, 53–66.

Kourilsky, M.L., & Graff, E. (1986). Children's use of cost-benefit analysis: Developmental or non-existent. In S. Hodkinson & D. Whitehead (Eds.), *Economic education: Research and development issues* (pp. 127–139). London: Longman, L.T.D.

Kourilsky, M.L., & Wittrock, M.C. (1987). Verbal and graphical strategies in the teaching of economics. *Teaching and Teacher Education*, *3*, 1–12.

Linden, M., & Wittrock, M.C. (1981). The teaching of reading comprehension according to the model of generative learning. *Reading Research Quarterly*, *17*, 44–57.

Macoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

MacKenzie, A.W., & White, R.T. (1982). Fieldwork in geography and long-term memory structures. *American Educational Research Journal*, *19*, 623–632.

Ortiz, E., & Kourilsky, M.L. (1985). The mini-society and mathematical reasoning: An exploratory study. *The Social Studies*, *76*, 69–75.

Osborne, R.J., & Wittrock, M.C. (1983). Learning science: A generative process. *Science Education*, *67*, 489–503.

Osborne, R.J., & Wittrock, M.C. (1985). The generative learning model and its implications for science education. *Studies in Science Education*, *12*, 59–87.

Peled, Z., & Wittrock, M.C. (1990). Generated meanings in the comprehension of word problems in mathematics. *Instructional Science*, *19*, 171–205.

Samuelson, P. (1983). *Economics from the heart: A Samuelson sampler*. [Preface]. New York: Harcourt & Brace.

Slavin, R.E. (1990). *Cooperative learning: Theory, research, and practice*. Englewood Cliffs, NJ: Prentice-Hall.

Walstad, W.B., & Soper, J.C. (1988). A report card on the economic literacy of U.S. high school students. *American Economic Review Proceedings*, *78*, 251–256.

Wittrock, M.C. (1974). Learning as a generative process. *Educational Psychologist*, *11*, 87–95.

Wittrock, M.C. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Wittrock, M.C. (1990). Generative processes of comprehension. *Educational Psychologist*, *24*, 345–376.

Wittrock, M.C. (1991). Generative teaching of comprehension. *Elementary School Journal*, *92*, 167–182.

Wittrock, M.C., & Alesandrini, K. (1990). Generation of summaries and analogies and analytic and holistic abilities. *American Educational Research Journal*, *27*, 489–502.

Wittrock, M.C., Marks, C.B., & Doctorow, M.J. (1975). Reading as a generative process. *Journal of Educational Psychology*, *67*, 484–489.

Part XV

MUSIC, ART

79

RESEARCH ON EXPERT PERFORMANCE AND DELIBERATE PRACTICE

Implications for the education of amateur musicians and music students

A. C. Lehmann and K. A. Ericsson

Following an overview of the current knowledge about the structure and acquisition of expert performance in the arts, sciences and sports, we discuss practical implications for music training, focussing on the development of levels of instrumental skill typically attained by high school students and amateurs. Recent studies found that even the highest levels of music achievement are primarily the result of skill acquisition and physiological adaptation in response to extended deliberate practice. Increases in performance over historical time also document the importance of training and practice. Although learning conditions encountered by music students and amateurs often may be less favorable than learning environments in which experts develop, the quality of training can be increased at all levels of performance by incorporating features commonly found in the training of experts (individualized practice assignments, improved monitoring of feedback).

During the last several decades there has been a growing interest in the study of complex everyday activities and their development. More specifically within cognitive psychology, scientists are studying the structure and the acquisition of expert performance in many domains of expertise including chess, medicine, and sports. One of the goals of this research was to better understand the development of high levels of performance in order to improve training for future experts as well as for amateurs and other individuals aspiring to more modest levels of mastery. One of the most extensively researched domains of expertise, besides chess, has been instrumental music performance.

Source: *Psychomusicology*, 1997, 16, 40–58.

The influence of practice, training and innate musical talent on performance has been examined several times in recent reviews (Ericsson, 1996; Ericsson & Charness, 1994; Ericsson, Krampe, & Tesch-Römer, 1993; Howe, Davidson, & Sloboda, 1998). The general conclusion from results obtained in many domains of expertise is that practice and training is important, that is, predictive of performance at all levels including the elite levels. Clearly, the highest levels of performance in a domain require optimal training conditions and learning environments. It is thus possible that deficiencies in training environment and learning resources account for some of the failures of even highly motivated individuals to reach the expert level. This paper intends to bridge the gap between the research done by psychologists and the implications that this research might have for the teaching and learning of music in instructional settings such as the music studio and classroom.

We will first explore why individual differences in observed performances are not always attributed to practice and training. Then we will describe aptitude testing in music as a logical consequence of the concept of innate talents. In the third section we will offer evidence for the importance of practice and training for attaining high levels of performance. The fourth section reviews different training activities and their effectiveness in skill development. For example, we describe active forms of deliberate learning which provide more effective methods for improving the structure of music performance than mindless repetition and drill. Finally, we discuss the application of expertise and practice research to learning situations of the studio and classroom.

The practitioner's dilemma of explaining skill differences

In the world of music and in many sports, teachers and students often believe that some people have "it" (whatever "it" is) and that some do not. According to this view, even large amounts of practice will not be sufficient to reach high levels of performance for those people who lack "it." The idea that abilities are genetically determined has a long tradition, and it has spawned psychometric attempts in every area of human performance, especially intelligence and related concepts. Later, we will show three separate lines of evidence that support the theory that high levels of music performance can be enhanced through practice, and that the notion of innate talent may not be necessary.

There are several reasons why beliefs in innate musical talent are so common among musicians. There is a historical tradition from as early as the Renaissance when artists claimed that God-given gifts made them predestined for artistic careers. Examples include geniuses such as Michelangelo and Leonardo da Vinci (Ericsson & Charness, 1994). Also, around the turn of the 19th century, when many musicians' biographies were written, philosophical ideas consistent with the talent concept made their way into the common understanding of creative and re-creative artistic behavior (e.g.,

Einstein, 1947; DeNora, 1995). Finally, music teachers in classrooms and studios try to provide their students with the same instruction and practice assignments, but they still observe large individual differences in their students' attained performance that they believe cannot be attributed to external training conditions.

Explanations of individual differences in performance often depend on the specific situation and also on our beliefs. When two students give a music performance, differences in their skills often are noticeable even to the untrained observer. An explanation based on acquired skills, practice, and training is more likely when certain cues are available, that is, when one of the players is older and has had several more years of training than the other. However, when two students taught by the same music teacher are similar in age and years of training, then it is difficult to explain observable differences in performance in terms of differences in external conditions. In this case, it would seem more likely that the individual differences are due to inborn capacities for music (innate talent). As the next example from the domain of sports demonstrates, the factors influencing observed differences in skill are sometimes very subtle and difficult to notice.

Boucher and Mutimer (1994) found that Canadian hockey players in the National Hockey League (NHL) were much more likely to be born in January-March than in October-December. The explanation for this surprising phenomenon relies on relative age effects and is rather simple. When children start playing hockey they are grouped according to age. Since the playing season starts in winter, the players born early in a given year would be many months older and have advantages in strength, motor skills, size, and experience over their younger teammates, who were born later in the same year.[1] The coach is likely to attribute some of the older players' better performance to innate talent and thus probably will give them more support, playing time, and better training opportunities. Whatever the selection mechanism may be, over time it favors those born in the first half of the year. A similar age effect was documented in tennis and soccer (Dudlink, 1994). Yet in soccer there is no size advantage for adult elite soccer players, and one study found that the better adult soccer players were shorter than other athletes (Medvet, 1966). In this case, the initial advantage of increased relative body size as a child eventually would turn into a disadvantage for the adult performer. The benefit of relative age effects are not linked to the beginning of the year (January through March), but rather coincide with the start of the cycles relevant for a certain domain, regardless where in the course of the year those cycles start and end (e.g., fall for academic activities).

Teachers who have to make judgments about the potential of certain students should be advised to carefully review the above criteria. Furthermore, knowledge about the child's family background, current developmental situation of a child, and hidden factors may well influence a decision with far-reaching educational consequences.

Aptitude testing as an alternative to simple attribution of innate talent

When music students have reached a sufficient level of skill it is possible to observe and evaluate their music performance as well as identify the students showing exceptional achievements and promise. However, it is far more difficult to uncover alleged music talent among children and adults who do not play an instrument. Although we now know that even early signs of "promise" are often disappointingly elusive (e.g., Howe, Davidson, Moore, & Sloboda, 1995), researchers in the field of musical talent long have tried to design psychometric tests of aptitude or potential for music training. Much of the current evidence cited in support of the validity of these aptitude tests comes from studies where large groups of children are tested in order to discover those individuals with "hidden talents." If these "talented" children were given appropriate training, presumably they would improve far more rapidly than other randomly selected children receiving the same training. However, this evidence may not necessarily show that the tests of music aptitude measure innate capacities and talents.

Most developers of aptitude tests claim that their tests can predict future performance (e.g., Gordon, 1967, but see Boyle, 1992, for a different view). However, research on prediction of performance in areas other than music has shown surprisingly low predictive validity of aptitude scores (such as intelligence) for many types of job performance after many years of work experience (e.g., Hulin, Henry, & Noon, 1990). Is it possible that music aptitude tests measure psychological or physiological characteristics of children lacking music training that may not have reliable consequences for future achievement in music? Let us consider the following example involving the seemingly obvious disability of being flat footed!

To assess deformities of the foot, orthopedists use an index called the arch index, which is derived by making a footprint and dividing the measured width of the heel area by the width of the arch area. Although the individual variability of this index in the general population is large (Staheli, Chew & Corbett, 1987), researchers found no evidence that the flexible flat foot in any variation produced disability in the absence of other clinical problems (Harris & Beath, 1948, cited in Staheli et al., 1987, p. 427). More directly relevant to music performance, Henson and Wyke (1982) showed that professional orchestral players scored higher than the average population on only three of the six subtests of Seashore's Measurement of Musical Talent, and even significantly worse than the average for the test of timbre. Thus, even professional musical achievements may not be meaningfully related to scores on music aptitude tests. In a recent review Lehmann (1997a) found that correlations between musical aptitude test scores and musical achievements are consistently low. The primary exception to this rule is mentioned in Gordon (1995), where the music aptitude test scores showed moderate

predictive validity after two and three years of music instruction. However, this study did not assess the prior music training of the students taking the music aptitude test. If high scoring students indeed had received more music instruction, and if formal instruction improved the test score–which is likely (Lehmann, 1997a)–then the high predictive validity of the aptitude test for future music performance may be confounded with the effects of formal instruction, or for that matter, the effects and informal musical training prior to the test. In the area of development of oral and written language, such informal training in the form of targeted parent activities, such as storybook reading and parent teaching about literacy, have been shown to influence young children's performances on standardized measures of achievement (Sénéchal, LeFevre, Thomas, & Daley, 1998). Similar effects may be observed from musical activity in the home of the students (cf. Bamberger, 1991). We know of no experimental studies on the effects of informal musical training on music aptitude scores.

In sum, although some music aptitude tests show a degree of predictive validity for the performance of traditional Western instruments, formal or informal training may possibly account for most of this predictive validity. In the absence of experimental studies we cannot clearly distinguish the relative contribution of innate aptitudes and acquired skills for the performance on the aptitude tests. Moreover, there may exist measurable individual differences among untrained children that lack significant consequences for future levels of music achievement. In the next section we review the evidence showing that performance can be enhanced greatly through practice and training independent from assumptions of aptitudes and innate talents.

Three types of evidence demonstrating effects of practice and training

Practice and experience are necessary for elite performance

Recent reviews show that extended engagement in performance activities is absolutely necessary to attain an expert level (Ericsson & Charness, 1994; Ericsson & Lehmann, 1996). Longitudinal studies have shown that performance does not increase in sudden jumps, but rather gradually. This implies that very high levels of performance are reached only through steady progress over many years of engagement in domain related activities. This also might apply to child prodigies, whose performance is vastly superior to that of their peers (see Wagner & Stanovich, 1996, p. 202, for a claim regarding reading achievements; see also Howe, 1990, for a general argument). Also, with maintained intense engagement in their field, expert performers continue to improve their performance beyond the age of physical maturation (the late teens in industrialized countries) for many years and even decades. In fact, the age at which performers typically reach their career peaks in the arts and sciences extends about two decades beyond their performance

maturation in the 30s and 40s (Ericsson, 1990; Lehman, 1953). This clearly implies that experience in one's field is absolutely necessary for individuals to improve their performance. The final and most compelling evidence for the necessity of vast experience prior to attaining high levels of performance is that even the most "talented" individuals in a wide range of sports, science, and arts require about ten years of intense involvement before they reach an international level (Simon & Chase, 1973; Ericsson et al., 1993). However, extensive experience, by itself, is not sufficient for attaining elite performance and the correlation between amount of mere experience and level of performance often is surprisingly low (Ericsson & Lehmann, 1996).

A closer association has been found between attained level of performance and a particular type of practice which Ericsson et al. (1993) called *deliberate practice* (practice activities involving specific goals and strategies). The investigators found a high correlation between indicators of attained performance of musicians and the amount of deliberate practice accumulated during their musical development. By age 20, the best violinists in their study had spent an average of over 10,000 hours in deliberate practice. This number was about 2,500 hours greater than the accumulated practice times of the violinists in the study's intermediate group and about 5,000 hours greater than the expert violinists in the least accomplished group. Sloboda, Davidson, Howe, and Moore (1996) showed that for music students, higher achieving students practiced significantly more than lower achieving students. Additional work has shown that the relation between deliberate practice and performance holds for domains of expertise other than music, such as individual and team sports and chess (Ericsson, 1996; Helsen, Starkes & Hodges, 1998).

Adaptations observed in experts

The second type of evidence for the acquired nature of abilities refers to observed physiological, cognitive and psychomotor characteristics that are associated with very high levels of performance. For a long time those characteristics were believed to reflect innate talent. A few of the many striking examples concern the larger hearts of elite endurance athletes and the larger number of capillaries supplying blood to their muscles (Ericsson & Lehmann, 1996). However, we now know that the vast majority of these anatomical and physiological characteristics are consequences of the intense training. On a small scale, we all experience physiological changes in response to engaging in everyday activities. For example, many days of yard work may bring about calluses and maybe even slightly firmer biceps. Of course, some of the physiological changes found in different domains of expertise, such as the larger hearts of athletes, have been shown to revert back to normal values when training ceases and there is no further demand for that type of extraordinary heart capacity.

Many thousands of hours of training also lead to measurable adaptations in musicians. For example, violinists and pianists maintain certain typical body positions when playing their instruments, and their respective abilities to rotate their forearms is modified in an instrument-specific way (Wagner, 1988). Violinists have a larger forearm supination (i.e., rotation so that the palm faces upward), while pianists have a larger forearm pronation (rotation so that the palm faces downward). Adaptations observed in experts even include very specialized structural changes in the brain. For example, the area of the cortex associated with control of the left hand fingers, especially the little finger, is enlarged for advanced stringplayers who started training at young ages (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1996). Also, practice related enlargements of the cortical areas activated by the presentation of complex tones have been found in musicians but not in nonmusicians. These and other results suggest a use-dependent functional reorganization in the sensory cortex (Pantev, Oostenveld, Engelien, et al., 1998). All those physiological characteristics that include changes to muscles, bones, and brain structure appear to be adaptations that experts acquire through extensive training in response to task demands.

There also are mental abilities that have been associated with high levels of performance, in particular extraordinary memory performance. Experts are known to deliberately increase their ability to plan and reason as well as their knowledge and the ability to access memory (Ericsson & Kintsch, 1995). In the process, and without training memory performance for its own sake, experts' domain-specific memory performance increases.[2] However, this memory advantage over less skilled individuals is restricted to meaningful stimuli in the domain and essentially disappears when experts are asked to recall random configurations of the same stimuli (for a review, see Ericsson & Lehmann, 1996). In summary, most of the physical and mental characteristic features of experts appear to be domain-specific adaptations to the typical demands, induced through practice and training.[3]

Historical increases

The final type of evidence showing how even high levels of performance can be increased further comes from historical changes in performance which can be demonstrated in music, but to date have been mainly documented in other domains of expertise such as sports and sciences. For example, Johnny Weissmuller (early "Tarzan" actor), won on Olympic gold medal in 1924 by swimming the 100 m freestyle in under 60 s. This world record was matched some 40 years later in 1964 by a woman (Dawn Fraser). Today, the same swimming time is no more than a very good time for a student athlete from a typical high school. Although the goal of musicians is not to play their musical instrument faster or louder than previously done, historical developments in skill akin to those in sports have been documented.

Recently we investigated historical changes in music performance by focusing on the complexity of performed music (Lehmann & Ericsson, 1998a). We correlated the dates of composition for the piano sonatas by Haydn, Clementi, Mozart, Beethoven, and Schubert, with contemporary complexity ratings published for those works. The results showed that sonatas from later periods tended to be rated more difficult than those from earlier ones. In a different analysis we documented an increase in the levels of achievement of child prodigies over a comparable time period. We collected information on the degree of precocity of piano prodigies from the last three centuries and found that more recent prodigies had played more difficult pieces at younger ages than their famous predecessors. Both these increases in the level of music performance can be explained by changes in training which have allowed later performers to achieve higher levels of technical performance faster than previous performers. In sports and music alike, the level of achievement that only a century ago was attributed by contemporaries to the unique innate talents possessed by a performer, is today regularly achieved by a large number of individuals after extended training. These previous levels of expert performance do not appear to require special innate talents, at least by today's standards, but they are viewed as predictable consequences of appropriate instruction and extended deliberate practice.

In conclusion, three types of evidence reveal the plasticity of human performance. The highest levels of performance are not fixed and immutable as shown by steady historical increases in elite performance. The level of attained performance rarely is constrained by anatomical, physiological or mental characteristics because those characteristics are shown to adapt and change in response to appropriate types of extended deliberate practice.

The development of skill and associated mental mechanisms through training

Although the importance of practice may seem obvious to many musicians, it is necessary to consider it more systematically. As mentioned earlier, even among highly skilled musicians there is a close relation between attained performance and the amount of deliberate practice they have accumulated throughout their musical development. Consequently, a more detailed analysis of the structure of practice is needed to identify the essential mechanisms of effective training. We will first describe the characteristics and constraints of deliberate practice and contrast it with mere experience and engagement in music related activities. In the final section we will discuss implications for learning in the classroom and studio.

Characteristics of deliberate practice

The theoretical concept of deliberate practice is restricted to learning activities with specific goals and activities. A student who engages in drill while thinking

about something else may experience only minor, if any, benefits for improvement in music performance. Ericsson, Krampe, and Tesch-Römer (1993) have defined deliberate practice as a structured activity designed to improve performance. Of course, this activity is embedded in the larger phenomenological context of practice. Deliberate practice has well-defined goals and the outcome is monitored carefully to see if the goals have been met. Although not all behaviors that a student displays during a practice session meet the requirements of deliberate practice, they may serve a multitude of different functions, such as maintaining interest and motivation through play and relaxation.

Some pedagogues do not acknowledge the effort and intentional nature of the development of expert skill which is inherent in the deliberate practice concept. For example, Kohut (1992) argues for a natural ability to learn which supposedly is lost during socialization in today's society, and which rests on two main learning mechanisms, namely trial and error and imitation. From our theoretical perspective, this approach, which parallels contemporary views of general education (see Ericsson, 1998, for a critical commentary), is only part of what constitutes musical learning. Although trial and error and imitation might feel "natural" and effortless, they are not always efficient processes. Imagine a child learning to drive a car by trial and error or yourself trying to pole vault by imitation. Mere experience and spontaneous engagement in activities do not automatically result in the types of complex skills that experts acquire. The complex skills of expert performers appear to require the intentional, designed activities of deliberate practice for their acquisition.

Factors constraining deliberate practice

Ericsson et al. (1993) identified several prerequisites for effective practice. First, an individual needs sufficient access to training facilities, appropriate training exercises and proper sequencing of instruction. Second, practice is an inherently effortful process that is limited by human attentional resources. We find that experts engage in deliberate practice for four to five hours on a daily basis. This seems to be the maximum that adult experts can maintain on an extended regular basis; children's daily practice times are even shorter. Finally, deliberate practice requires sustained concentration and effort and consequently differs form many other similar activities that are more inherently enjoyable, such as playful interaction with peers. Next we will turn to the question of motivation in deliberate practice.

Some critics of the theoretical framework of deliberate practice (e.g., Gardner, 1995; Sternberg, 1996; Winner, 1996) point out that not all children attain the same final level of performance even when they are given the same instruction and opportunities for deliberate practice. However, actual engagement in deliberate practice requires an active act on the part of the

child to concentrate on the task and to monitor the performance; consequently motivation becomes a key constraint (Ericsson et al., 1993; Ericsson & Charness, 1994; 1995). Critics of the expert performance framework do not disagree with the close link between motivation and high levels of performance. However, most advocates of innate talent believe that the talented individual is motivated to practice because he or she can perform tasks better than the less talented individual. In contrast, proponents for the expert performance framework point out that deliberate practice requires individuals to set performance goals beyond their current level of achievement, thus leading to repeated failures until eventual mastery is achieved (Ericsson et al., 1993; Ericsson, 1998). Given that at least some of the factors determining the level of motivation are due to prior experiences and environmental influences, it is essential that teachers and home environments be examined carefully to establish their role in the acquisition of high levels of performance.

Some studies attest to the importance of first teachers. The best teachers for beginners, especially every young ones, need to know the children's personality well enough to keep them motivated to maintain practice by showing enthusiasm and ample encouragement. Therefore, it is not surprising to find that the first teachers of prodigious piano players most often were members of the prodigy's household (Lehmann, 1997a). Later teachers promote skill development more rigorously and rely more on the student's intrinsic motivation (L'Hommedieu, 1992).

Experience and practice

As mentioned earlier, deliberate practice is different from mere experience. This distinction has become clear in research on expertise, where highly paid experts, such as medical experts, despite their large amounts of experience do not necessarily outperform less experienced professionals in the same domain on routine tasks. Related findings exist for those domains in which decision making is the crucial component, such as in selecting stocks (Ericsson & Lehmann, 1996). One likely factor limiting learning and improvement of performance in many professional settings is that immediate feedback on performance is typically not available. In the absence of guiding feedback, improvement of accuracy of performance might be nearly impossible. Although mere repetition of the same performance will not lead to better performance per se, it may lead to increased automaticity and reduced effort (Shiffrin & Schneider, 1977).

Let us, for example, consider sight-reading and piano accompaniment that involve performance of unfamiliar music after only limited rehearsal. We found that the time that pianists had engaged in regular sight-reading or accompanying activities was predictive of their sight-reading performance (Lehmann & Ericsson, 1993; 1996). Individuals who had accumulated more

hours of regular sight-reading performed more accurately than the others. We also found that pianists' accompanying repertoire predicted sight-reading ability over and above experience. The accompanying repertoire contained pieces that the pianists had specifically rehearsed for accompanying purposes, and these pieces typically included accompaniments to string and wind sonatas and to operas and oratorios. Together, these results demonstrate well the interplay of experience and self-imposed challenge. Consider the church musician who sight-reads hymns for many years. Once the required level of mastery has been reached we would not expect this musician to improve beyond this level unless the complexity of the material is deliberately increased. In contrast, educational environments are typically designed to continuously challenge the students' level of achievement. In our sample of college age pianists, their accompanying activities forced them to constantly encounter and master new and more difficult pieces, thus confounding experience and practice opportunities. Also, they imposed challenges on themselves by specifically learning certain materials. Without knowledge about these additional challenges, we would overestimate the contribution of mere experience to improvements of performance. Put simply, the best sight-readers were those who challenged themselves and learned complex repertoire in addition to their regular sight-reading engagements in choir, church, or chamber music.

Emerging mental representations as a result of practice

Building mastery in a domain and finding the least effortful method to attain a specific performance goal are very different activities. This distinction is crucial for separating a performance that has been entrenched through mindless drill from one that is flexible and adaptable through the use of mental representations—a hallmark of expert performance (Ericsson & Lehmann, 1996; Ericsson, 1997; Lehmann 1997b). Everyday observations support this distinction. For example, a person can learn to pronounce perfectly a number of useful conversational phrases in a foreign language by rote and appear proficient when asking for the menu in a restaurant. However, when the waiter attempts to initiate a conversation, the real limits of language comprehension and production become evident. Similarly, some music students can play a few well-entrenched pieces from start to finish but cannot improvise, sight-read, or efficiently learn new pieces; whereas others who can play the same level pieces are also able to display all other aspects of musicianship. Thus, superficially similar performances, in this case the performance of a given piece, may be mediated by distinctly different underlying mechanisms.

To explain those contrasting mechanisms of performance, we have proposed a model of mental representation that implicates complex mental processes rather than assuming that rehearsed music is played in an automated fashion without conscious control by the musician. Three different types of mental representations are deemed necessary for expert musicians

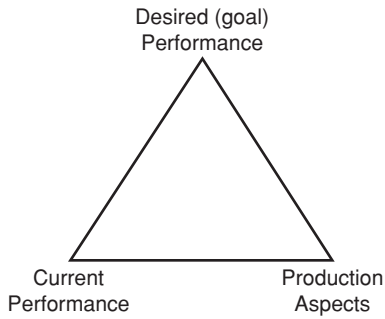


Figure 1 Mental representations necessary for expert performance in music

(see Figure 1). The first mental representation is that of the desired (goal) performance, which contains the musician's representation of how the piece should sound. The second representation is related to the musician's ability to implement the goal representation, and this production representation includes the knowledge of and control over the instrument. A third representation contains the current performance, and this representation is related to the musician's ability to monitor his or her own performance. These representations are interconnected but can be experimentally separated and studied. Think-aloud protocols (verbal reports) are an important data source in this process. In a laboratory experiment where subjects memorized a short piece of music and then reproduced it under changed performance demands from memory, we found evidence for the cognitive mediation of music performance. The verbal reports revealed the different mental representations, and subjects even described representational errors such as playing right hand notes with the left hand when they were asked to only perform the left hand (Lehmann & Ericsson, 1997).

Implications of research on expertise and deliberate practice for music teaching

Can we apply what we know about expert levels of performance to the acquisition of skills at more modest levels of performance? To understand how such findings could be extrapolated to a typical classroom or studio situation, we must consider the differences between learning environments of expert performers and more common learning environments (see Table 1). The characteristic that most prominently distinguishes between the skill acquisition of experts and public music education is that group instruction prevails in the latter while one-on-one tutoring is the norm for the training of experts. Also, the early start of training of high achieving children contrasts with the relatively late start of instruction in public music education. Finally, the training of future expert performers is focused on an ultimate

Table 1 A comparison of learning environments of prodigies, confirmed experts, amateur musicians, and music students

Typical learning environments encountered by amateur musicians and music students	Typical learning environments encountered by prodigies and confirmed experts
Group instruction	Individual (one-on-one tutoring)
Relatively late start (sometime during schooling)	Very early start (usually before formal schooling)
Short training period with short-term goals	Extended training periods with long-term goals

performance years away, whereas music schools are evaluated primarily by the success of public performances weeks or months away. However, we believe that by providing more supervised practice, more accurate feedback, varied training activities, and a clear goal as to what the final level of performance should be, group settings can provide some of the benefits that experts receive from their training environments.

Supervised practice

In contrast to many sports in which the coach is often present during practice, music students typically retreat to the practice room and work by themselves. This also is true for public school situations in which students are encouraged to practice their parts at home. However, a recent study of past piano prodigies showed that virtually all the child pianists had engaged in practice under supervision of an adult (Lehmann, 1997a). An informal survey of biographies of contemporary prodigies including Yo-Yo Ma, Cecilia Bartoli, Evgeny Kissin, and Sarah Chang reveals similar patterns. It is probably not the musical ability of the supervising persons that is critical, but their aid in maintaining the child's concentration (on-task behavior) and monitoring as well as their occasional suggestions for improvement. In schools, improved supervision should be possible by encouraging students to practice in pairs. Some anecdotal evidence from teachers who advocate "practice partner" activities suggests that improved performance and motivation result. However, we are not aware of controlled experimental study evaluating those or similar claims. Some studio teachers, especially those teaching in the Suzuki tradition, include the parents in the practice process; others make their students keep a practice diary. The former may actually satisfy some prerequisites of deliberate practice, while the latter simply ensures that some practice time is spent at the instrument, whether or not this time is filled with beneficial activities. Systematic empirical research evaluating these suggestions is necessary before firm recommendations can be made.

Goal setting and feedback

Given that appropriate feedback, goal setting and monitoring have been proposed as key characteristics of effective practice (see Ericsson et al., 1993; Singer, Murphey & Tennant, 1993, for a review based on studies in sports), providing goals and feedback also would be expected to be beneficial in the music classroom. However, these methods are far more difficult to implement for a group situation than in one-on-one instruction, where an individual performer's specific goal can be related to a specific outcome. For example, although the band director may verbally provide a goal for the group and then give a summary feedback, verbal instructions and feedback may not be given concerning those aspects to which a particular student was actually attending. Instead, students should be involved in the process of goal setting and monitoring and internalize this process. Some teachers have their students evaluate each others' performance. In addition to the disciplinary benefit of this "keep everybody busy" method, it might also serve an important function in allowing the students to improve their abilities to internalize goal setting and to monitor outcomes. Also, many method books provide checklists to be used for setting the right goals or more generally for monitoring progress during practice. Any pedagogical and technical device that supports more specific goal setting and subsequent monitoring will improve the quality of practice in accordance with the deliberate practice concept.

Multiple training activities

Deliberate practice in music typically refer to individuals' solitary efforts to improve a particular aspect of their performance. However, the concept of deliberate practice includes any training activity for which goals have been defined and feedback is available. Each learning activity in turn promotes the acquisition of an associated skill; all these skills together lead to a structure which supports a particular performance. To become creative improvisers, jazz musicians imitate models, listen to recordings, and try to understand the style of a given performer. Chess experts spend large amounts of time studying published chess games by masters, predicting the next move and then comparing their predicted move to what the master actually did. Discrepancies between a chosen chess move and the master's move then are analyzed. Assuming that the master's move was indeed the best choice, this activity combines goal setting and instant feedback.

From the chess and jazz example it becomes clear that training activities of experts are closely matched to task demands of the domain, and students should use those same or adapted activities to start developing similar skill and associated underlying representations. For example, expert wind players may work out the peculiarities of their instrument (e.g., natural pitch tendencies), and all students should be taught to work on these performance

aspects. Generally, only the careful study of experts' training activities will allow us to translate and adapt some of them to group settings which often include novices and more advanced music students.

Awareness of mental representations

Performance is primarily constrained by the requirements of the task and environmental expectations. For example, since work and play do not typically require individuals to exhibit their maximal performance, professionals are often able to increase their performance through training when given external incentives to do so (see Ericsson et al., 1993). Also, plateaus in improvement of performance are mostly due to insufficient motivation and rewards to keep expending the time and effort to improve; but in some cases progress is arrested due to strategies that cannot be extended easily to higher levels of performance. Therefore, music educators should carefully determine the ultimate performance goal for each student and design an educational plan to teach the mental representations that will be necessary to reach that goal.

The ability to perform a specific piece of music can be attained by many different methods. Some methods provide short cuts to minimize effort, while others help to develop and refine complex mental representations. For example, rote memorization is an efficient process for memorizing short easy pieces, but this type of memorization may prove ineffective for longer pieces with a more complex structure or when the memorized material needs to be manipulated during performance. For improvement to continue, strategies that have been useful in the early stages of skill acquisition, such as rote memorization, may need to be replaced later by more effective ones. For example, professional actors have developed strategies for understanding the fictional character they portray on stage so that they can reproduce their lines and actions to virtually eliminate the need for rote memorization (Noice & Noice, 1997). Musicians also show complex memorization strategies (Chaffin & Imreh, 1997; Hallam, 1997; Lehmann & Ericsson, 1998b). Thus, encouraging young musicians to memorize their music by rote might in fact be counterproductive because it may prevent—or at least discourage—the use of higher level musical representations. A similar argument could be made for rote imitation, which may well be an excellent behavior in acquiring musical interpretations at lower levels of performance with sometimes astonishing results; but it is unlikely to suffice when the student attempts to develop into a mature artist. Anecdotal evidence suggest that the ability to imitate by rote versus a more complex artistic skill distinguishes prodigies who eventually fail from those who succeed (e.g., Cortot, 1935).[4]

It should be stressed that many of the above thoughts and suggestions are not new and most of them can be found in teaching methods and treatises. However, while our theoretical framework allows us to explicate why these

things work, most master teachers use them intuitively without reflecting on the corresponding learning mechanisms.

Conclusion and discussion

Skills can be fostered and developed, and training and practice play a crucial part in this process. As the discussion of practice and acquired mental representations has conveyed, there are, however, often qualitative differences in the underlying mechanisms that mediate seemingly similar levels of performance. This may be especially true at lower levels of performance. Mere repetition and experience lead to more fluent performance, but by themselves do not lead to the mental representations that experts employ (e.g., the difference between rote memorization and more complex internal representations of a piece of music that allow experts to adapt to different performance problems). The close association between extended training and performance with the associated specific physiological, psychomotor, and cognitive adaptations provides strong evidence for the acquired nature of skills. Thus, explaining high level performance solely in terms of innate talent might mislead parents and teachers to settle for short-term successes rather than to support and foster the covert and (admittedly) slow emergence of superior skills and representations.

The most effective activities for improving performance are effortful and involve conscious decisions with trade-offs and life-long consequences. They draw heavily upon a person's motivational resources, but not for mindless perserverance to repeat a section 1,000 times, but for motivation to concentrate and deliberately build an integrated skill. Simply playing the same problem section correctly several times or slowing down the tempo may eliminate an immediate performance problem, but it may be far less useful in the long run than carefully studying the cause of the problem (and finding the solution). This procedure would remedy a particular deficiency (as well as similar future problems) by building appropriate mental representations. Even for a student not aspiring to be a virtuoso, the realization of this interplay of final goal and often short-term effort would be of motivational and educational value.

Would it be possible for educators to identify the types of representations a student would most benefit from at a given level of performance? Could educators then design instruction that would enable development of the cognitive and psychomotor skills necessary for expert performance? We believe that further advances in our understanding of music performance learning will depend greatly on future studies of the mental representations that experts are able to develop in relatively optimal learning environments. Not until we understand how these representations can be acquired reliably under optimal conditions can we seriously discuss potential implications for public music education. Regardless, we believe that increased insights into

the general processes and cognitive mechanisms of effective deliberate learning will help both teachers and students to reach their immediate performance goals in a manner that is consistent with their long-term goals as musicians. Accordingly, the often-encountered emphasis on short-term performance goals without regard for the mental processes that mediate the attained performance may well be short sighted and might indirectly limit many children's ultimate level of music achievement.

Author note

Parts of this paper were presented by the first author as an invited paper at the 1997 SUNCOAST Music Education Forum in Tampa, Florida. We also wish to thank R. Woody, L. Hill, and two anonymous reviewers for their comments on earlier versions of this paper.

Notes

1 For example, the height of the tallest 4-year-olds (those in the 95th percentile) is comparable to that of the average 5-year-olds (those in the 50th percentile, Malina & Bouchard, 1991, p. 50). Thus, especially at younger ages, height differences are considerable among children who are almost 12 months apart.
2 This is anecdotal musical evidence and includes young W. A. Mozart's alleged transcription of a piece after a "single" hearing (MGG, 1961, vol. 9, 701; see Stafford, 1991, for a more moderate version of the same anecdote), and the ability of conductors (e.g., George Szell and Arturo Toscanini) to perform even long operas without a score.
3 Thus far the only exception to this general rule is height, for which current research shows a strong genetic determination (Ericsson & Lehmann, 1996).
4 Alfred Cortot is one of the most famous piano teachers of the 20th century.

References

Bamberger, J. (1991). *The mind behind the musical ear: How children develop musical intelligence*. Cambridge, MA: Harvard University Press.

Boucher, J. L., & Mutimer, B. T. (1994). The relative age phenomenon in sport: A replication and extension with ice-hockey players. *Research Quarterly for Exercise and Sport*, *65*, 377–381.

Boyle, D. (1992). Evaluation of music ability. In R. Colwell (Ed.), *Handbook of research on music teaching and learning* (pp. 247–265). New York: Schirmer.

Chaffin, R., & Imreh, G. (1997). "Pulling teeth and torture": Musical memory and problem solving. *Thinking and Reasoning*, *3*(4), 315–336.

Cortot, A. (1935). Do infant prodigies become great musicians? *Music and Letters*, *16*, 124–128.

DeNora, T. (1995). *Beethoven and the construction of genius: Musical politics in Vienna, 1792–1803*. Berkeley, CA: University of California Press.

Dudlink, A. (1994, April 14). Birth date and sporting success. *Nature*, *368*, 592.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1996). Increased cortical representation of the fingers of the left hand in string players. *Science*, *268*, 111–114.

Einstein, A. (1947). *Music in the Romantic era*. New York: Norton.

Ericsson, K. A. (1990). Peak performance and age: An examination of peak performance in sports. In P. Baltes & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 164–195). Cambridge, UK: Cambridge University Press.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 1–50). Mahwah, NJ: Erlbaum.

Ericsson, K. A. (1997). Deliberate practice and the acquisition of expert performance: An overview. In H. Jörgensen & A. C. Lehmann (Eds.), *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 9–51). Oslo, Norway: Norges Musikkhogskole.

Ericsson, K. A. (1998). [Commentary on J. R. Anderson's, L. Reder's and H. A. Simon's paper "Radical constructivism, mathematics education and cognitive psychology".] In D. Ravitch (Ed.), *Brookings Papers on Educational Policy 1998*, (pp. 255–264). Washington, DC: Brookings Institution Press.

Ericsson, K. A., & Charness, N. (1994). Expert performance. Its structure and acquisition. *American Psychologist*, *49*, 725–747.

Ericsson, K. A., & Charness, N. (1995). Abilities: Evidence for talent or characteristics acquired through engagement in relevant activities? [Reply to Gardnor, 1995]. *American Psychologist*, *50*, 803–804.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence for maximal adaptations to task constraints. *Annual Review of Psychology*, *47*, 273–305.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406.

Gardner, H. (1995). Why would anyone become an expert? [Commentary on Ericsson & Charness, 1994]. *American Psychologist*, *50*, 802–803.

Gordon, E. E. (1967). *A three-year longitudinal predictive validity study of the musical aptitude profile*. Iowa City: University of Iowa Press.

Gordon, E. E. (1995). *Musical Aptitude Profile (Manual)*. Chicago, IL: GIA.

Hallam, S. (1997). The development of memorisation strategies in musicians: Implications for education. *British Journal of Music Education*, *14*(1), 87–97.

Helson, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport and Exercise Psychology*, *20*, 12–34.

Henson, R. A., & Wyke, M. A. (1982). The performance of professional musicians on the Seashore measures of musical talent: An unexpected finding. *Cortex*, *18*, 153–158.

Howe, M. J. A. (1990). *The origins of exceptional abilities*. Oxford, UK: Blackwell.

Howe, M. J. A., Davidson, J. W., Moore, D. J., & Sloboda, J. A. (1995). Are there early childhood signs of musical ability? *Psychology of Music*, *23*, 162–176.

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: reality or myth. *Behavioral and Brain Sciences*, *21*, 399–407.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin*, *107*, 328–340.

Kohut, D. L. (1992). *Musical performance: Learning theory and pedagogy*. Champaign, IL: Stipes.

L'hommedieu, R. L. (1992). The management of selected educational process variables. *Dissertation Abstracts International*, *43–06A*, 1836.

Lehman, H. C. (1953). Age and achievement. Princeton, NJ: Princeton University Press.

Lehmann, A. C., & Ericsson, K. A. (1998a). The historical development of domains of expertise: Performance standards and innovations in music. In A. Steptoe (ed.), *Genius and the mind: Studies of creativity and temperament in the historical record* (pp. 67–94). Oxford: Oxford University Press.

Lehmann, A. C., & Ericsson, K. A. (1998b). Preparation of a public piano performance: The relation between practice and performance. *Musicae Scientiae*, *2*, 69–94.

Lehmann, A. C. (1997a). Acquisition of expertise in music: Efficiency of deliberate practice as a moderating variable in accounting for sub-expert performance. In I. Deliege & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 165–191). London: Erlbaum (UK), Taylor & Francis.

Lehmann, A. C. (1997b). Acquired mental representations in music performance: Anecdotal and preliminary empirical evidence. In H. Jörgensen & A. C. Lehmann (Eds.), *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 141–164). Oslo, Norway: Norges Musikkhogskole.

Lehmann, A. C., & Ericsson, K. A. (1993). Sight-reading ability of expert pianists in the context of piano accompanying. *Psychomusicology*, *12*, 182–195.

Lehmann, A. C., & Ericsson, K. A. (1996). Structure and acquisition of expert accompanying and sight-reading performance. *Psychomusicology*, *15*, 1–29.

Lehmann, A. C., & Ericsson, K. A. (1997). Expert pianists' mental representations: Evidence from successful adaptation to unexpected performance demands. In A. Gabrielsson (Ed.), *Proceedings of the 3rd Triennial ESCOM Conference* (pp. 165–169). Uppsala, Sweden: Uppsala University.

Malina, R. M., & Bouchard, C. (1991). *Growth, maturation, and physical activity*. Champaign, IL: Human Kinetics.

Medvet, R. (1966). Body height and predisposition for certain sports. *Journal of Sports Medicine and Physical Fitness*, *6*(2), 89–91.

MGG (1961). Wolfgang Amadeus Mozart. In F. Blume (Ed.), *Musik in Geschichte und Gegenwart (Vol. 9)*. Kassel, Germany: Bärenreiter (Reprint).

Noice, T., & Noice, H. (1997). The nature of expertise in professional acting. Mahwah, NJ: LEA.

Pantev, C., Oostenveld, R., Engelien, A., Ross, V., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811–814.

Sénéchal, M., LeFevre, J. A., Thomas, E. M., Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly*, *13*, 96–116.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Simon, H. A., & W. G. Chase (1973). Skill in chess. *American Scientist*, *61*, 394–403.
Singer, R. N., Murphey, M., & Tennant, L. K. (Eds.), *Handbook of research on sport psychology*. New York: Macmillam.
Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. (1996). The role of practice in the development of expert musical performance. *British Journal of Psychology*, *87*, 287–309.
Stafford, W. (1991). *The Mozart myths: A critical reassessment*. Stanford, CA: Stanford University Press.
Staheli, L. T., Chew, D. E., & Corbett, M. (1987). The longitudinal arch. *Journal of Bone and Joint Surgery*, *69A*(3), 426–428.
Sternberg, R. J. (1996). Costs of expertise. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 347–354). Mahwah, NJ: Erlbaum.
Wagner, C. (1988). The pianist's hand: anthropometry and biomechanics. *Ergonomics*, *31*, 97–131.
Wagner, R. K., & Stanovich, K. E. (1996). Expertise in reading. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 189–226). Mahwah, NJ: Erlbaum.
Winner, E. (1996). The rage to master: The decisive role of talent in the visual arts. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 271–302). Mahwah, NJ: Erlbaum.

80

HOW CAN CHINESE CHILDREN DRAW SO WELL?

E. Winner

Chinese children do not draw childish drawings. Young children in China make drawings that seem to challenge theories of the developmental course of drawing skill (e.g., Gardner, 1980; Kellogg, 1969; Winner, 1982). Instead of the large, messy, semi-expressionist paintings seen in American preschools and elementary schools, in which children reveal their own invented ways of representing, one sees in China small, neat paintings in which children display their precocious ability to master adult ways of representing the world.

Chinese children learn to paint in two styles that at first glance appear very different—"Western-style" watercolors and traditional Chinese ink paintings. Western-style paintings are typically colorful scenes in which the entire space is filled with small figures engaged in a dazzling variety of activities—walking, sitting, jumping, shown from the back and looking up, running with one leg going back and in foreshortening, holding umbrellas, etc. (figs. 1 and 2). These postures (never seen in drawings by such young children in the West) are no mean feat to depict, and it is quite breathtaking to encounter such a repertoire in children even as young as six. The figures depicted are almost always children, and they all look the same—a round head, two big black dots for eyes, and a smiling mouth. The brush is used with great skill: the figures are outlined in black, and the colors are bright. For some colors the brush is applied directly from the paint jar, while for others the paint is thinned with water to create a translucent effect.

The Chinese ink paintings are in the suggestive style of traditional Chinese art. Often they consist of a single figure well positioned amidst a sea of blank space. They depict traditional Chinese subjects—shrimp, goldfish, monkeys, roosters, chickens, bamboo, a branch with a flower—all painted in a way that (at least to my relatively untutored eye) is nearly indistinguishable from these same subjects painted by the ancient masters (figs. 3 and 4).

Source: *Journal of Aesthetic Education*, 1989, 23(1), 41–63.

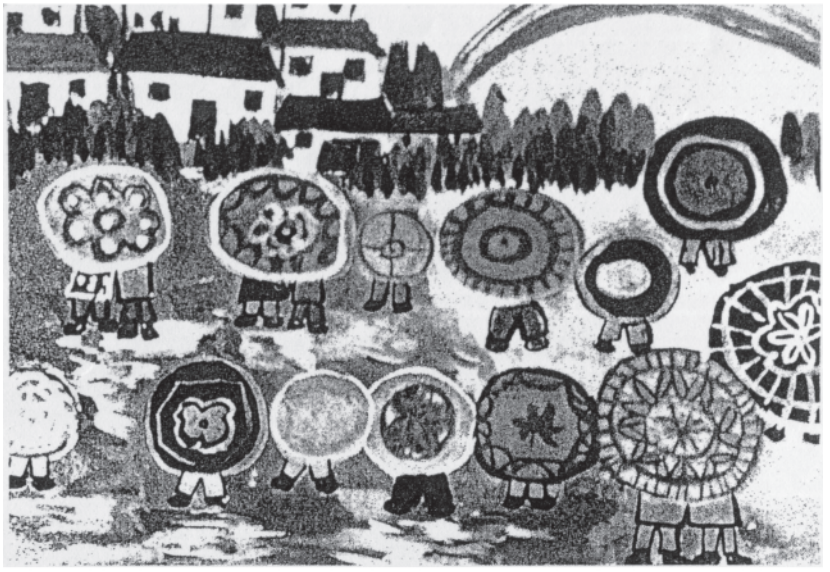


Figure 1 Children holding umbrellas: Prize-winning Western-style painting by a nine-year-old

Figure 2 Six-year-old's Western-style painting showing perspective

Figure 3 Traditional Chinese flower painting by six-year-old

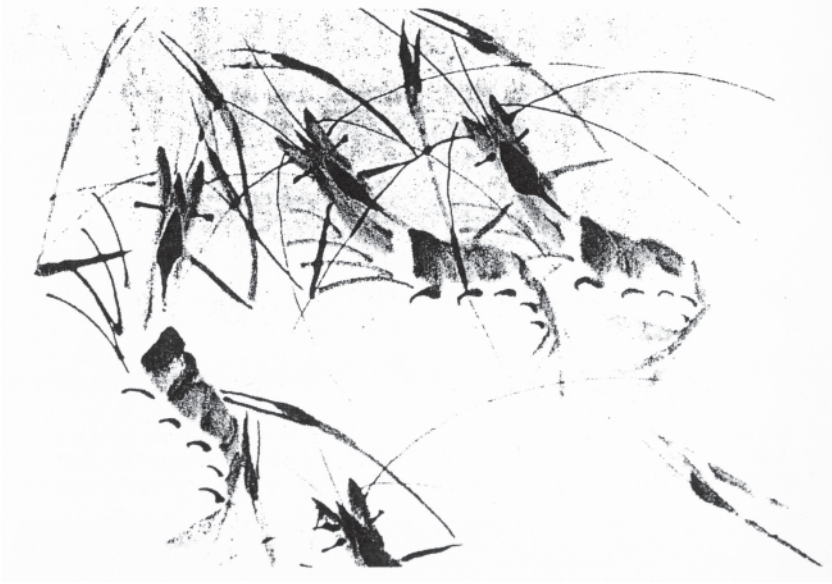


Figure 4 Traditional Chinese shrimp painting by eight-year-old

Such adultlike paintings can be seen in endless supply in any good urban school in China. It is not just the gifted children who draw in this way. While the artistically talented children excel and while it is *their* pictures that are selected for international exhibitions, the paintings of the average children are also very skilled relative to those of Western children.

In the spring of 1987, I spent two months in China observing Chinese art classes. Two questions guided my observations. I wanted to understand the teaching methods used to inculcate the high level of drawing skill displayed by even ordinary children. And second, I wanted to understand the relationship between skill training and creativity. Was technical skill taught at the expense of creativity, imagination, innovation? Or did the acquisition of skills and schemas allow children to be creative by freeing them from the *struggle* of trying to make things "look right"?

Teaching the child who is in control

Within a week of my arrival, I realized that I could only hope to answer these questions by looking at Chinese art education within the context of Chinese childrearing. Chinese children behave very differently from American children, not only in art class, but in most other situations as well. In China, toddlers and babies are very quiet. I saw three- and four-year-olds eating quietly and neatly with their parents at restaurants and sitting motionless at an adult concert. And I saw little children riding on the handlebars of a parent's bicycle, a practice that could never be carried out if the children did not sit absolutely still. I came to suspect that Chinese children learn to draw with such skill at least in part because of their willingness to *comply* and their ability to *concentrate*. These two proclivities are developed in children by methods of childrearing very different from those used in the West.

According to my informants, the childrearing practices that create quiet, well-behaved children have remained relatively unaltered over the years. Molding the child to be passive begins at birth. Although the practice is dying out among the educated, most babies are still bound so that they cannot move for the first one to three months of life. Even after the binding is taken off, babies are held constantly. And they are bundled up with so many layers of clothing (to protect against catching cold) that they can hardly move.

Toilet training by conditioning begins almost at birth. In this, as in so many ways, the child is molded as quickly as possible into an adult. The procedure is quite simple. The parent holds the baby over a potty and whistles. This procedure is repeated every twenty minutes until the baby urinates. Eventually the child comes to associate the whistle with urination, and the whistle is enough to elicit urination. By one year or earlier, children are completely toilet trained.

Adults often tease children, and I came to see this as a way of showing the child who is in control. An American woman living in Beijing with her

husband and one-year-old baby daughter told me how Chinese parents routinely hold spoons teasingly just out of reach of the baby. The baby tries to grab the spoon, but the adult never lets the baby win. Finally, frustrated, the baby stops reaching. Then, when reaching behavior has been properly "extinguished," the parent teaches the child how to hold a spoon. The message is clear: Do it when *we* think you are ready. And do it *our* way. As I learned, the same message is conveyed in the art room.

Thus, Chinese children are discouraged from "natural" attempts at independence; and, as they are held all of the time, dependency needs are immediately met. The result—the quiet, well-controlled children seen everywhere: in buses, on bicycles, in concerts, in restaurants, and in the classroom.

Discipline in the kindergarten

The early and successful inculcation of compliance, carried out by the family, is continued by the kindergarten teacher. Even three-year-olds can sit still, pay attention, and focus on a task for up to thirty minutes at a time. Never once did I see children refuse to do a task, ask to go to the bathroom in the middle of a lesson, or announce that they were "done" before the bell rang.

Classes are very "teacher centered." All eyes are on the teacher at all times. This contrasts strikingly with the American classroom, where students often work together and teach themselves, calling on the teacher only when they need help. The Chinese teacher often asks questions, but the questions are not what we think of as questions. The questions asked are always ones for which the children know the answers, and the teacher knows that the answers are known. "What song did we learn last time?" "What kind of animal is this?" (as the teacher holds up a picture of a rabbit). "What kinds of animals live in the forest?" "What is the title of this song?"

Children either raise their hands to answer these questions or, in response to some subtle cue from the teacher, they chant out the answers in unison. Answering questions in Chinese schools is a performance, much like a ritualized dance. The concept (advocated only in the best Western schools) of posing questions that puzzle, perturb, or elicit many kinds of answers, even wrong ones, seems to be lacking. Children are set up to succeed; they are not prodded into thinking, questioning, or wondering. The same philosophy permeates the art room: children are not challenged to think visually and to solve visual problems; instead, they are given solutions in forms that are easy to master, and they are expected to master them. For instance, children are not ever expected to figure out by themselves how to draw something new; instead, they are shown how to draw images step by step, line by line.

It is not surprising, then, that the children themselves never ask questions. There is little need to ask a question when one is presented with clear material with no puzzles. Teachers explain everything and leave nothing for the child to ask about. I heard only one question in two months of

classroom observations. A primary school child raised her hand after the art teacher had explained the drawing instructions to ask, "Are we allowed to add a tree?" (The answer was "Yes.") The same kind of reluctance to ask questions has often been noted by Americans observing or teaching Chinese students at the university level.

Paradoxically, the Chinese instill a competitive spirit in their students along with compliance and passivity. Beginning in preschool, children receive evaluations. I often saw posters on the classroom wall listing each student's name, with red flowers stamped after some names. Children could get a stamp in various categories, including obedience, care for property, and attendance. The fact that the evaluations were posted publicly makes for considerable competition. Children even compete to get red stamps for cooperation! This emphasis on competition, which has its roots in the ancient imperial exam system where ranking was all important, is to be found in art classes as well as in academic subjects and classroom behavior. In one kindergarten, I saw children's work displaved in a hall exhibit with a poster on which was written the question, "Who draws the best?"

Here, then, is the context in which art education takes place: placid, controllable, unquestioning children expecting to be led step by step; the desire to meet the teacher's expectations so that one can receive a high evaluation or even win an art competition; quiet, ordered, teacher-centered classrooms; and parents who want their only child to excel in any classroom subject, including the visual arts.

A look behind the scenes

The Chinese educational system is governed by a uniform curriculum and national textbooks which all teachers, even art teachers, must use. (Recently, I have learned that there are efforts underway to allow teachers the choice of several possible textbooks and curricula.) Textbooks in art contain the lessons deemed appropriate for each age. Thus, when I describe a particular class, the observations can be generalized to other classes (at least in urban schools) all over China. Indeed, we often saw the same lesson repeated almost verbatim in two different cities.

There is also a uniformity in teaching methods across the various forms of visual arts. In effect the medium makes no difference. The age-old method for teaching calligraphy provides the standard; teaching methods in the various art forms are slight modifications of the techniques used to teach calligraphy.

Learning the painstaking art of calligraphy

Calligraphy training begins in first grade for forty minutes a week. Children learn how to sit, how to hold the brush for the different kinds of strokes, how to prepare the ink, and how to mix the ink with water to achieve

precisely the right quality of tone required for the different brush strokes. The goal is to master the tradition, not to go beyond it. Students are not expected to discover how to use the brush to create various effects, but rather to learn what the masters have already discovered.

In a fourth grade calligraphy class, I watched children prepare their ink by rubbing an ink stone into a small well at their desks. Each child had a textbook containing rows of Chinese characters. The characters were written on top of grids, which made it easy to see the underlying structural skeleton of the characters. Under each character was drawn the same character, but this time only with thin lines. Students first traced over the fully formed characters in the top row, and then filled out the lines in the lower characters so that the brush strokes were of the appropriate thickness and tone.

The teacher began the class by drawing a character, using white chalk on the blackboard. Next, she held up a large version of the same character, painted in black ink on white rice paper. This character was also on the page open in each child's text. The teacher asked the children about the composition of the character. Some characters have a "top-bottom" structure, while others have a left-right structure. The character with which the lesson began was a top-bottom one in which the ratio of top to bottom was one to three. Children were questioned about the ratio and were asked to name characters with the same kind of structure. This was excellent training in seeing composition, but was not apparently viewed as an aesthetic task. Rather, it was seen as the first step in teaching the child how to paint a new character. The teacher went on to hold up various characters and to ask for their names and a description of their ratios.

The teacher then demonstrated the procedure for making a character by painting on the blackboard with a thick brush dipped in white paint. As she drew each line, she told the class how to hold the brush, demonstrating how to use more force at the ends of lines. She also demonstrated how *not* to make a line—by applying uniform pressure and achieving a line uniform in thickness. She showed how to start each line slowly, with force, and then to pick up speed and use less pressure. She then pointed to the different strokes she had made. Each stroke had a name, and the children chanted out in unison the name of each stroke.

The children were then ready to begin tracing the same characters in their texts. All of the students sat up straight and waited for the teacher to tell them to begin. They went to work on signal and worked slowly and painstakingly, in total silence. They spent an average of two or three minutes tracing each character. The work was handed in at the end of class, to be returned later with a numerical grade. Corrections would be made by the teacher in red ink.

These children had been practicing calligraphy in this method since the first grade. According to the teacher, children do not rebel against this and do not question why they must learn this ancient technique, which is no longer in daily use (but of course remains as an art form). No child seemed

bored or irritated by the exacting nature of the task. In fact, as in all of the classes I observed, the students seemed engaged and showed an impressive level of concentration.

I was struck by the extent to which painting (and even sculpture and handicrafts) was taught like calligraphy class: models are provided by the teacher and the text book. These models are painstakingly copied by the students, and it is clear to everybody that there are right and wrong ways to draw—just as there are correct and incorrect ways to hold the calligraphy brush and mix the ink. The notions of art as process, as visual problem solving, or as innovation are conspicuously absent.

Practicing the traditional schemas of Chinese painting

In the elite kindergartens and primary schools, children are taught traditional methods of Chinese ink-and-brush painting. In a fifth grade painting class, fifty students sat quietly at small desks lined up in rows. On each desk was a long sheet of paper (about two-and-a-half feet long, one foot wide). The sheet covered the desk, and the rest of the sheet was rolled up under the desk. Space limitations rendered it impossible for children to see the whole sheet all at once.

The teacher began the class by lecturing to the students on the kind of painting style that they would learn that day. He described the style as very simple and suggestive, consisting of a few well-chosen lines. He demonstrated it by painting a fish with a few swift lines on a sheet of paper tacked on the board. He then hung up a famous painting of chickens by a great master done in this style. He pointed out the large amount of blank space, the lack of realism, the attempt to suggest rather than to depict in detail. The assignment for the day was to paint chickens in this style and then to add one or two bunches of grapes to the picture—a traditional motif in Chinese painting.

The teacher then took down the master's painting and demonstrated step by step how to draw chickens and grapes in this style. He took out two brushes and demonstrated how to dip them in water and wipe the water off onto special paper. He made a very black mark and asked what color it was. Students chanted out, in unison, "very dark." He then painted the chickens, stroke by stroke, asking the students to count the number of strokes that formed each part. He pointed out that he was using the soft brush. Then he took the other brush and told the class that this one was for making watery light lines rather than the dark ones.

After he had painted two chickens in the style he had introduced, the teacher drew on the board chalk sketches of chickens engaged in various activities—fighting, eating, and running. The students were told that they could draw any of these.

Then the teacher moved on to demonstrate the strokes for painting grapes. Here he introduced yet another technique: the tip of the finger is used to

make a small round shape. Students were told that they should make some grapes darker than others. When asked the reason for this, students immediately said that this was in order to make some look nearer (the dark ones) and some farther away (the light ones). Clearly they had been taught this rule. This is in essence atmospheric perspective, except that the differences in distance between the grapes are tiny. When one looks at a bunch of grapes, the near grapes do not really appear darker; whereas when one looks at a vista of mountains, the ones in the distance really do appear lighter. Thus, the students were learning a color perspective that is not true to life but is rather a code, one used in traditional Chinese painting.

The students were then instructed to make a painting of chickens and grapes in the demonstrated style. The students were expected to use the models provided by the teacher at the front of the class and also the models in their textbooks, which were identical to the ones the teacher had demonstrated. The textbook makes things easy for the teacher as well as the student, and if necessary the teacher can be but one schema (or formula) ahead of the student. The students worked with great concentration for the half hour that remained after the fifteen minutes of lecture-demonstration.

Although the similarities between teaching calligraphy and Chinese painting are evident, children are allowed more leeway in painting than in calligraphy. In the class just described, students were allowed to draw any of the chickens the teacher had sketched in chalk, but only one of these had also been shown in Chinese brush style. Thus they were expected to do a bit more than direct copying. Moreover, they were not told how many chickens to make, nor what positions to put them in. Students are expected to master a rich array of schemas but are then allowed to arrange these elements somewhat as they wish. For example, one child tried to connect the chickens and grapes by painting the chickens standing under the grapes, some of which were falling down into the chickens' open beaks. The teacher praised this work for its originality. However, no child altered the basic chicken or grape schema. This would have been viewed as incorrect rather than original.

Chinese painting is taught in kindergarten as well as in elementary school. I was struck by how brilliantly a class of six-year-olds had mastered paintings of shrimp, crabs, and fish. I wondered whether they had only learned how to apply this style in painting the figures they had been taught or whether they had actually mastered a style of painting that would generalize to new subjects.

I conducted an informal experiment to answer this question. I got the answer—by and large, they had not mastered a generative style—and in the process discovered something else as well. I asked the children to draw our baby stroller—an object they had never seen before and hence an object for which they had not been taught a formulaic Chinese painting schema. I asked the children to paint this object in the style in which they had learned to paint shrimp and crabs.

The most interesting aspect of this little experiment was the reaction of teachers and administrators. They were horrified when they saw what I intended to ask the children to paint. They tried to talk me out of it, suggesting instead that I pick out a stuffed animal out of the toy room, something with which the children would be familiar. They insisted that this task was too difficult. Children would not know what the stroller was and hence would not be motivated to draw it. The teachers were concerned that children might fail. They were not used to giving children challenges to solve. Instead, the method used in China is to teach in incremental and imitative fashion, so that even the ordinary or slow child will succeed.

The teachers had underestimated what their children could do. The children studied the stroller carefully and made paintings that were detailed and realistic rather than in the impressionistic style of Chinese painting. Thus, they did not generalize the traditional style to a new subject—the answer to my initial question. Most had not learned a generative style, but instead had mastered a set of rules for painting shrimp, another set for goldfish, yet another for crabs, and so forth.

Although this finding suggests a limitation to the kind of schema training going on, the drawings produced also revealed a dramatic payoff of the educational method. The children produced drawings far more skilled and realistic than their American peers would have created. Thus, the intensive practice in seeing, and in eye-hand coordination, that these children receive from copying pictures seems to generalize to drawing from life.

Accumulating Western schemas

Western painting is defined as any kind of painting or drawing in which children use Western materials: pencils, craypas, markers, or watercolors on nonporous paper. In Chinese painting lessons, children learn to paint traditional schemas used by the ancient masters; in Western painting lessons, children learn to paint cartoonlike schemas borrowed from Western cartoons. These cartoon schemas are now seen in Chinese children's comics and newspapers (figs. 5 and 6) and also often decorate classroom walls. A comparison of figures 5 and 1 reveals the striking similarity between the adult and child schemas for drawing figures with umbrellas; a comparison of figures 6 and 2 reveals the similarity between adult and child schemas for drawing children.

Despite the superficial differences between them, Western painting is taught identically to Chinese painting: the teacher and the textbook provide the schemas; children are expected to master a wide variety of schemas and are allowed some flexibility in how they combine these elements into a finished work. At the heart of each lesson is the mastery of one or more schemas or formulae for drawing particular objects.

In a primary school drawing class, I watched first graders learn how to draw penguins. The class began with a color videotape of penguins in their

Figure 5 Figures holding umbrella: Adult-drawn image from children's newspaper

natural habitat. Children were then asked to do a one-minute sketch of the penguin's environment (e.g., ice, snow, water, rocks) as background for their pictures. Then the "real" class began. "Today," the teacher said, "you will learn how to draw a penguin." She showed how to block out the oblong shape of the body. "First make two points on your paper showing how tall it should be, then two more for the width, then connect the four points with a line." Next she showed how to make the head, and how *not* to make it. Figure 7a shows the correct way: here the head fuses into the body; figure 7b shows the wrong way: here the head is a circle stuck onto the body. Next to this incorrect sketch she drew a big X to remind children not to draw it in this way. Children were then told to locate the position of the eye. The teacher drew the eye, then the wing and feet. Next, she showed how to color in the black parts. The drawing at the bottom of figure 7 is the teacher's completed model. Children also had their textbooks open to the penguin lesson. The book provided a step-by-step method for drawing a penguin, identical to the one that the teacher had demonstrated.

The children then went to work. They drew on small paper, with pencils and thin markers. They copied the model that the teacher had provided, although the background was drawn from imagination. Even the background was constrained, however. One child drew flowers and grass and was chided by the teacher and told to erase them. "Use your head. There are no flowers at the South Pole." On the other hand, one child asked the teacher to show him how to make a penguin diving, and she told him to try it by himself and to use his imagination. Thus, children were given conflicting messages: although they were told to draw the penguin according to the model, they were also told to use

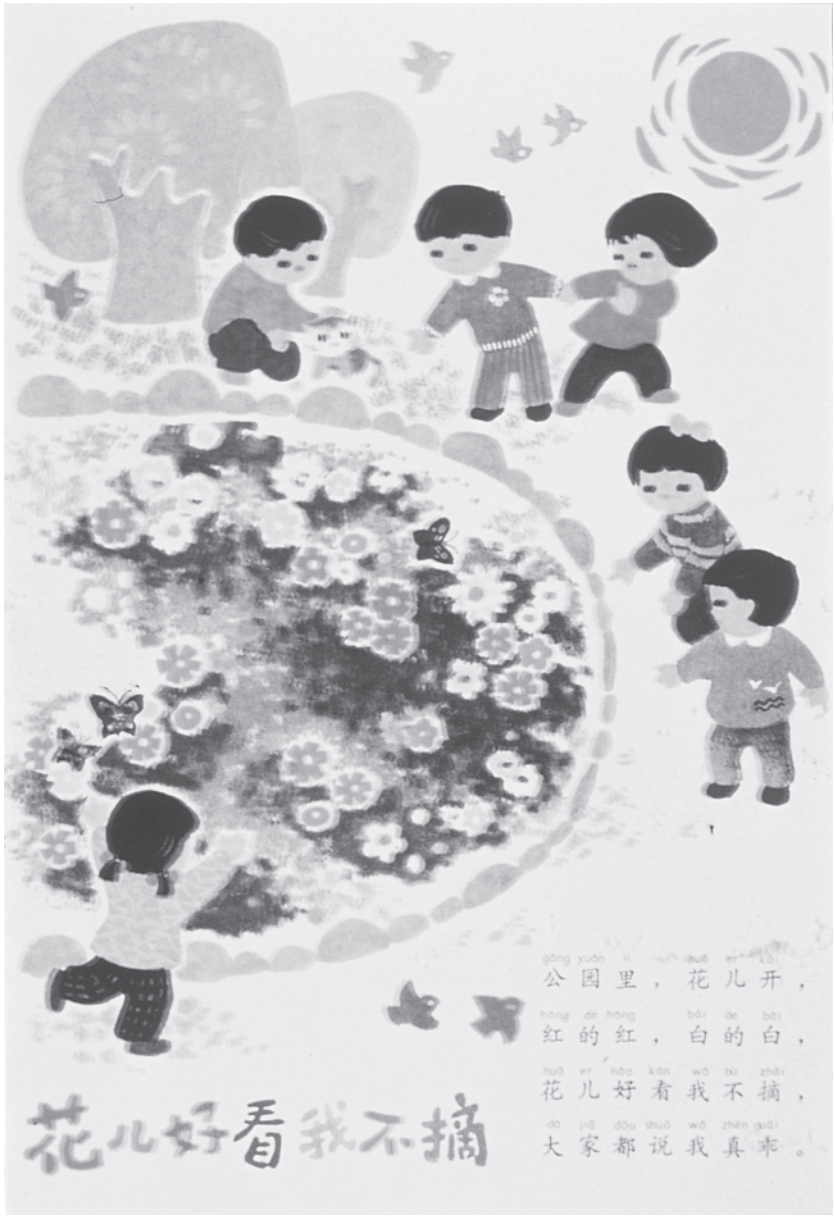


Figure 6 Children playing: Adult-drawn image from children's newspaper

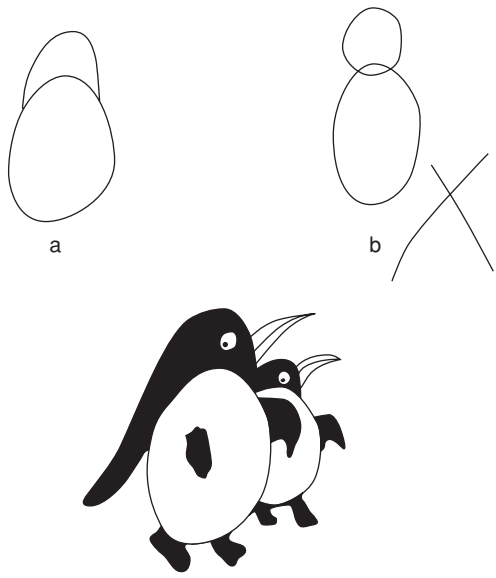


Figure 7 (a) Correct penguin schema demonstrated by teacher; (b) incorrect penguin schema demonstrated by teacher; (bottom) teacher's completed penguin schema from which children were to copy

their imagination; and although they were told to draw the background from their imagination, some children discovered that certain contents were proscribed.

As in the chicken-grape lesson, the teacher's schemas were taken right out of the children's textbooks. But children did not only draw from the models provided—they also brought their own. I noticed two children copying from scraps of paper on which were detailed, skilled pencil drawings of penguins that looked much like the ones in the texts. I asked one of these children what she was copying from, and she replied, with no embarrassment, that her father had drawn this to help her. This child had capitalized on the fact that she had been told ahead of time that the penguin lesson was coming up.

The images on the videotape were not mentioned once the tape had been turned off; hence, the tape remained disconnected from the rest of the lesson. The children copied only from the drawn models. Thus, children learn to draw from drawings rather than from life. The lessons are not tailored to teach children how to see the three-dimensional world; rather, they teach children two-dimensional formulae.

What the arts should be for

Societies have viewed the value of the visual arts in at least four different ways. Art has been valued as an embodiment of beauty; as a conveyer of

moral and political values; as a means of emotional expression; and, more rarely, as an enterprise involving the intellect in special ways. In China, one hears a great deal about the first two uses of art, but nothing about the latter two.

Beauty

The Chinese believe that beauty is central to the arts. The term "aesthetic education" translates as "beauty education." As discussed by Gardner (this issue), artworks are esteemed if they are beautiful and looked down upon if they are not. There are rules for achieving beauty that govern subject matter, composition, color, use of brush, and so forth. The notion that an artist would want to paint squalor, use muddy or clashing colors, or create an unbalanced composition is incomprehensible. Art teachers and even artists found it difficult to understand that in the West art is valued if it offers a new vision, whether or not the painting produced is beautiful.

Morality

Along with the traditional emphasis on beauty there is a link between beauty and goodness. This link is an ancient one, although the link to specific political precepts is a socialist invention. The arts are seen explicitly as a means of moral education. It is believed that art education should function to inculcate "the morally correct idea of beauty." Art is not to be enjoyed or produced for its own sake, but must have representational content that serves moral and political ends. Thus, it is not surprising that nonrepresentational art is shunned and that socialist realism prevails in adult, Western-style art.

Nonrepresentational art is not only avoided, it may also not even be recognized. I will cite one example of such failure of recognition which may well be due to teachers' lack of exposure to abstract art. In a kindergarten collage class, children were given colored circles and asked to use them to make a picture of something from their daily lives. Most children made neat representational scenes. But two made what looked to me like abstract designs. I mentioned this to the teacher, but she felt that one picture was a flower and the other was a tablecloth. Of course a tablecloth can have a design on it, but the picture was seen as a representation of a tablecloth and not a mere abstract design.

Emotion

Western justifications of art as a form of self-expression were lacking. Westerners commonly believe that artists paint to express themselves and to work out their feelings. It is not unusual in the West to see children making drawings of things that have happened to them that have high affective

content—a visit to the doctor, the death of a pet, and the like. I did not find art put to such uses in China. The artistic tradition is highly formalized, and it is difficult to think of a bird-and-flower painting as something that could have served to express the artist's personal conception of the world. Similarly for child art. I never saw pictures that had a personal voice. Instead, they were exceedingly stylized. When I asked children if they ever felt the desire to draw something that they had just seen on the street or that had just happened to them, they looked at me blankly and said no. What they draw are schemas that they have learned in school.

In the West it is often believed that drawings can inform about the artist's personality or at least his personal vision. I cannot imagine anyone even raising the question in China since the art produced, both by children and adult artists, is so highly stylized.

Cognition

The view of the arts as cognitive, as activities that involve reflection, problem solving, and problem finding, was also absent. The concept of "visual thinking" (Arnheim, 1969) was difficult to explain to art teachers, possibly even to artists. This would probably be true in any culture in which art is highly formalized. If to be an artist is to master a craft, then there is no need for visual thinking because there are in effect no new problems to solve.

What should art education do?

In the West, influenced by the writings of Dewey (1933) and Piaget (1970), we value childhood as a special time that should not be rushed. Children are believed to have their own understanding of the world, and this understanding (although "wrong" by adult standards) has its own logic. Children, we believe, should be allowed to see the world in their own way. The aim of the educator is not to mold children in the image of the adult, but rather to pose challenging problems so that children will eventually discover for themselves more cognitively advanced ways of understanding.

In American art classes at their best, teachers do not give children the answers, but rather let them try to solve problems for themselves (including visual ones). So, for example, in an art class for kindergarten children, the Western teacher never explains *how* to draw something and rarely even suggests *what* should be drawn. When I explained this to Chinese teachers, they worried that the classroom would be chaotic, that children would never "learn," and that the teacher would have no role.

Chinese art education is more aptly described as art *training*. By this I mean that it is training in skill mastery rather than education in seeing and in solving visual problems. Two aspects about this training strike me as key: (1) the value placed on neatness and uniformity rather than on deviation

and creativity; (2) the value placed on schema mastery rather than on training the eye to break away from schemas.

Neatness and uniformity vs. deviation and creativity

Children are extremely neat in art class. The materials used are rarely messy ones. For the most part, children use ink or watercolors with thin brushes, pencil, craypas, and small pieces of paper. Children do not use easels, but work seated at their desks. The paintings that result are also very neat.

Given the teaching methods, it is not surprising to see a high degree of uniformity in the art produced. I often saw bulletin boards with children's drawings tacked up, and all the drawings looked the same—even in kindergarten, where we are so used to seeing diversity. It was not unusual to see thirty pictures done by one class that were all variations on a chicken schema or a goldfish formula.

In line with the emphasis on uniformity, one finds a lack of alternatives from which children can choose. In the United States, children have an excessive amount of free choice. There are innumerable after-school activities from which to choose. And if children tire of one, they are likely to switch to another the next semester. Not so in China. We visited a primary school in which children choose in first grade which after-school activity they will pursue—a musical instrument, dance, singing, calligraphy, or a visual art form (Western paining, Chinese painting, or handicrafts). They will then stick to this choice throughout the six years of elementary school. There is essentially no way to change.

Such a policy would be difficult to implement in the West, where freedom of choice is so highly valued. And of course any choice made in the first grade is likely to be made by parents, not children. When I asked whether children ever tried to change, or wanted to change, I was told that this rarely if ever has happened. Moreover, as is so often the case in China, the children in these extracurricular activities seemed thoroughly engaged and on their way to attaining high levels of skill within their often rather narrow areas of choice. Thus, they may spend six years mastering calligraphy or Chinese brush painting. In the United States, in contrast, children have so many alternatives from which to choose that they often fail to stick to anything long enough to gain mastery. My initial shock at such a forced marriage gave way to an appreciation of the advantages of this system.

Because of the emphasis on uniformity, there is little stress on creativity and deviation. In China, the act of painting is like performing a piece of music written by someone else. Contemporary artists often paint schemas invented centuries ago. Although artists may ultimately put their personal stamp on what they paint, this is analogous to putting one's own interpretation into a piece of music that one performs, rather than composing one's own piece.

For the Chinese, creativity seems to mean a slight departure, whereas in the West in means a sharp departure. The Chinese teacher's concept of a lesson in creative drawing is simply to allow children to draw from memory rather than from a model. What is never given is a task with visual challenges and no obvious solutions (e.g, draw a sad picture of a tree; draw a picture that looks heavy; make a collage with a surprising composition). Tasks are either rigidly structured or totally loose. But even the loose ones are graded on neatness and realism.

The teachers all insisted that they aim to teach not only basic skills (translate that as formula mastery) but also creativity and imagination. I asked the teacher of the penguin class what would count as creativity in the penguin task. She replied that if a child drew something not mentioned by the teacher or not included in the text, this would count as creative—for instance, if the child added an observation station, or a red flag, or showed the penguin in a different position. Recall the child in the Chinese painting class praised for creativity because he painted the chickens eating the falling grapes. Note that these are all representational deviations—new ways of combining old schemas—rather than stylistic ones such as alterations of the form of the schemas themselves.

Schemas: to be mastered or overcome?

Art education in China aims to give the child mastery of a rich repertoire of representational schemas. Visual exploration and invention have little place in the curriculum.

Look back for a moment at figures 1, 2, 3, and 4—the intricate Western-style watercolors and the simpler, impressionistic Chinese ink paintings. Different as they may seem on the surface, these two types of painting are in fact both constructed out of combinations of simple schemas that children have learned in incremental fashion in art class. To be sure, the schemas are possessed in abundance, and children are able to use them generatively. But the schemas themselves are passed down rather than created.

In my view, the best kind of art education in the West is based on principles diametrically opposed to Chinese principles. The aim is to stimulate children to break away from formulae rather than to perfect these formulae. Although no art can break away fully from pictorial schemas (cf. Gombrich, 1960), American art education strives to teach the eye to see beyond schemas. The best art teachers have invented exercises in which formulae will not work. Examples include contour drawing, drawing the negative rather than the positive space, copying a picture upside down so that one does not know what one is copying and thus cannot use a schema, and drawing a scene viewed through a frame that allows only part of a picture to be seen. One need only look at a book about techniques of teaching drawing to find many more examples (cf. Edwards, 1979). What is prized in the West is the

ability to see in a new, fresh way. If one paints according to old schemas, one has not seen in a new way, nor will one cause one's audience to see anew. Whether the art produced is pretty, or neat, or realistic is unimportant. If children's observational skills have been sharpened, if they have been made aware of design elements in their work, and if they are able to reflect about the visual decisions that they make, then Western art educators have achieved their goals.

This difference (schema mastery vs. novelty of vision) can be traced to the differences in the aesthetic traditions of East and West. Chinese artists were traditionally trained to master schemas. For example, in the seventeenth century *Mustard Seed Garden Manual of Painting*, stroke-by-stroke instructions are given for painting orchids, rocks, trees, people, etc., and rules of composition and color are explicitly stated (Sze, 1959). While this manual was criticized by some because it was seen as simply a set of instructions for beginners, the manual was very widely distributed and used (Sze, 1959). Moreover, this manual reflects the aesthetic ideals of the great Chinese painters.

Traditionally, students were supposed to master the art of their teacher and of the ancient masters. Only after the artist is advanced in age and experience might he be allowed to put his personal stamp on the traditional subjects such as birds and flowers. The goal is to master a tradition, not to start a new one.

Design and abstraction

Along with the representational goal of teaching the eye to see and the hand to draw "what is really there," American art education also seeks to sensitize children to aesthetic principles. These principles—design, color, texture, style, expression, etc.—have nothing to do with representation. To this end, American teachers offer exercises in design. For instance, students might be asked to make a picture composed of only ten squares, arranged in a surprising composition. Or students might be asked to make two paintings of the same subject, one in primary and one in secondary colors, and to then reflect about the expressive or compositional effects of these two different color choices.

One can understand the point of these types of design problems only if one has a sense of what abstract art is. Design exercises are ways of getting the student to think about the nonrepresentational components of art. Most Chinese art teachers and artists have had little exposure to abstract art and hence were puzzled by the exercises I described. Of course, the end product of these exercises might not look beautiful, but it is the process that matters (e.g., learning about design) and not only the product. For the Chinese, what is important is the product and the process of performing and executing it. For the Americans, what is important for the most part is not the product but the process of thinking, problem solving, or problem finding that goes into the product.

These different emphases need to be seen in the context of the different aesthetic traditions in China and the West. Chinese painting is highly stylized, and tradition is prized over radical novelty. Thus, if one wants to train children to paint in the traditional Chinese manner, an educational method that fosters the mastery of schemas is appropriate. Western art is much less stylized, and abstraction and novelty of vision are valued. Thus, an educational method that seeks to free children from schemas and sensitize them to design may be the more appropriate approach.

The Chinese insist that one must have skill before one can be creative. Yet preschoolers in the West belie this claim. They have little skill and, in my view, much creativity. Surely adult artists need both, and the appropriate time to get basic skills would seem to be in primary and high school. This is when children are ready and want to learn techniques and rules. But need there be a dichotomy? One can teach technique with exercises that call for creativity. For example, one can teach perspective by supplying a problem (e.g., draw two figures, one near and one far away) and allowing the child to invent the solution by herself. In my view, the kind of skill teaching used in China makes creativity unlikely to occur because children are given solutions instead of being asked to invent them.

In sum, Chinese art education does not train children to look at aesthetic aspects such as composition, style, color, or texture. Never did I hear those principles of design discussed as a teacher critiqued a student's work or instructed the class in how to proceed. Western art educators believe that students should learn to see such aspects not only in art but also in the environment. In China, there is no use of the environment for aesthetic training. Children are not taught to look around them, to notice the colors, shapes, and compositions of the landscape. It is as if art is disconnected from everything else: art is an escape from poverty and disorder into order and beauty.

What the West can learn from the Chinese example

I found much to admire in the Chinese art classroom, despite the fact that I was critical of the failure to teach children to solve visual problems in new ways. What impressed me the most was the high skill levels attained and the children's ability to concentrate and work until they had mastered a task. There is undoubtedly a benefit in learning to do something very well, no matter what it is that is learned. I have compared Chinese art education to Western art education *at its best*. Too often, however, Western art education is playtime with no inherent structure. Children are expected to be sloppy, and they are. Children are allowed to switch focus or give up on a task at whim. Should it surprise us, then, that Western adolescents are often alienated and bored, rather than engaged in some activity?

Western art educators might profit from the Chinese example without copying it. The Western classroom would benefit, I believe, from the reintroduction

of discipline—but discipline of a different sort than the Chinese use. Instead of imitation and drill, discipline ought to consist of working with care and concentration, sticking to a task over an extended period of time until a work feels as good as it can, and reflecting upon the problems one is trying to solve. Children in the West might then gain something of the engagement and technical proficiency that are so impressive in Chinese children while retaining the ability to discover their own solutions.

Acknowledgments

The conduct of this research and the preparation of this essay were made possible by support from the Rockefeller Brothers Fund to Harvard Project Zero and the Center for U.S.-China Arts Exchange of Columbia University. I am grateful to Rudolf Arnheim for comments on an earlier draft of this article.

References

Arnheim, R. (1969). *Visual thinking*. Berkeley: University of California Press.

Dewey, J. (1933). *How we think*. New York: Henry Regnery Company.

Edwards, B. (1979). *Drawing on the right side of the brain*. Los Angeles: Tarcher.

Gardner, H. (1980). *Artful scribbles: The significance of children's drawings*. New York: Basic.

Gombrich, E. (1960). *Art and illusion: A study of the psychology of pictorial representation*. Princeton, N.J.: Princeton University Press.

Kellogg, R. (1969). *Analyzing children's art*. Palo Alto, Calif.: National Press.

Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (vol. 1). New York: Wiley.

Sze, Mai-mai (1959). *The way of Chinese painting: Its ideas and technique*. With selections from the seventeenth-century *Mustard seed garden manual of painting*. New York: Random House.

Winner, E. (1982). *Invented worlds: The psychology of the arts*. Cambridge, Mass.: Harvard University Press.

Part XVI

SECOND LANGUAGE LEARNING

81

BILINGUALISM AND EDUCATION

K. Hakuta and E. E. Garcia

The concept of bilingualism as applied to individual children and to educational programs is discussed, and the history of research on bilingual children and bilingual education programs in the United States is reviewed. Bilingualism has been defined predominantly in linguistic dimensions despite the fact that bilingualism is correlated with a number of nonlinguistic social parameters. The linguistic handle has served policymakers well in focusing on an educationally vulnerable population of students, but the handle is inadequate as the single focus of educational intervention. Future research will have to be directed toward a multifaceted vison of bilingualism as a phenomenon embedded in society.

Bilingualism is a term that has been used to describe an attribute of individual children as well as social institutions. At both levels, the topic has been dominated by controversy. On the individual level, debate has centered on the possible costs and benefits of bilingualism in young children. On the societal level, fiery argument can be witnessed in the United States about the wisdom of bilingual education and the official support of languages other than English in public institutions. Particularly in the latter case, emotions run hot because of the symbolism contained in language and its correlation with ethnic group membership.

The controversy surrounding bilingualism is magnified by a sense of urgency generated by the changing demographic picture. In the United States, there are over 30 million individuals for whom English is not the primary language of the home. Of those, 2.5 million are children in the school age range, with this number expected to double by the year 2000. There are now many states in which the linguistic-minority school population is approaching 25% or more (Arizona, California, Colorado, Florida, New Mexico, New York, and Texas), and in many large urban school districts throughout the United States, 50% of the students may come from non-English-speaking homes.

Source: *American Psychologist*, 1989, 44(2), 374–379.

Whether the debate is over the merits of bilingualism in individuals or institutions, there is considerable confusion over a basic definitional issue. The problem can be succinctly stated as follows: Is bilingualism strictly the knowledge and usage of two linguistic systems, or does it involve the social dimensions encompassed by the languages? Oscillation between these linguistic and social perspectives on bilingualism has frequently led to misconceptions about the development of bilingual children as well as misunderstanding in educational initiatives to serve linguistic-minority populations.

As a case in point, consider the linguistic and social complexities contained in the following statement about school experiences by a ninth-grade Mexican-born boy who had immigrated from Mexico six months earlier:

> There is so much discrimination and hate. Even from other kids from Mexico who have been here longer. They don't treat us like brothers. They hate even more. It makes them feel more like natives. They want to be American. They don't want to speak Spanish to us, they already know English and how to act. If they're with us, other people will treat them more like wet backs, so they try to avoid us. (Olsen, 1988, p. 36)

Bilingualism, thought of simply as a bivariate function of linguistic proficiency in two languages, underrepresents the intricacies of the social setting. The history of research on bilingual children contains many false inferences about the effects of bilingualism based on a miscalculation of the complexity of the phenomenon. Similarly, current research to evaluate bilingual education programs takes an extremely narrow definition of bilingualism, that is, as the usage of two languages in instruction.

The importance of language in helping us understand the phenomenon is obvious. Nevertheless, language's accessibility to scientists must not be confused with its role in either the cause of problems or solutions to them. Wage distribution can be useful in telling us about the structure of racial discrimination, but changing wage distribution may not help solve the root causes of the problem. In a similar way, looking at language, we realize, only helps to facilitate the identification of problems and potential solutions, but additional steps are needed to provide adequate education to linguistic-minority students.

In this article we argue that although language provides an important empirical handle on the problems associated with bilingualism, one must be careful not to overattribute the causes of those problems to linguistic parameters. We provide brief overviews of the knowledge of bilingual children and bilingual education programs that has been gained through reliance on narrow linguistic definitions, bearing in mind its heuristic value. We there offer future directions for research.

The bilingual child

In the calculus of mental energy, what are the costs of bilingualism? Early research on the effects of bilingualism on immigrant children, conducted primarily at the turn of the century, painted a bleak picture. As Thompson (1952) wrote in summarizing this body of literature, "There can be no doubt that the child reared in a bilingual environment is handicapped in his language growth. One can debate the issue as to whether speech facility in two languages is worth the consequent retardation in the common language of the realm" (p. 367).

Much of this early work on bilingualism in children can be interpreted within the context of the social history surrounding the debate over the changing nature of imnigration in the early 1900s. The basic data to be explained were bilingual children's poor performances on various standardized tests of intelligence. From the empiricist point of view, the bilingualism of the children was thought to be a mental burden that caused lower levels of intelligence. This viewpoint was offered as an alternative to the hereditarian position, argued forcefully by prominent nativists such as Carl Brigham, Lewis Terman, and Florence Goodenough, that the new immigrants were simply from inferior genetic stock (Hakuta, 1986). Subscribers to the latter viewpoint sounded the social alarm that "these immigrants are beaten men from beaten races, representing the worst failures in the struggle for existence. . . . Europe is allowing its slums and its most stagnant reservoirs of degraded peasantry to be drained off upon our soil" (Francis Walker, quoted in Ayres, 1909, p. 103).

What is interesting about this early literature is its definition of bilingualism. The bilingual children included in these studies were not chosen on the basis of their linguistic abilities in the two languages. Rather, societal level criteria having to do with immigrant status were used, such as having a foreign last name (see Diaz, 1983). It is not clear whether the "bilingual" children in these studies were at all bilingual in their home language and English. Yet, on the basis of such studies using social rather than linguistic criteria, conclusions were drawn as to the effects of linguistic variables on intelligence. The point here is that language is a salient characteristic of children from immigrant and minority backgrounds that provides an opportune dumping ground for developmental problems that may or may not be related to language.

Research in the last few decades, fortunately, has developed considerable sophistication in understanding second-language acquisition and the nature of bilingualism. What has emerged is a relatively consistent set of answers to some fundamental questions about the linguistic and cognitive development of bilingual children. These answers argue against the early view—still held to be fact by some laypersons and educators—that bilingualism could be harmful to the child's mental development and that the native

language should be eliminated as quickly as possible if these effects are to be avoided.

Indeed, more recent studies suggest that all other things being equal, higher degrees of bilingualism are associated with higher levels of cognitive attainment (Diaz, 1983). Measures have included cognitive flexibility, metalinguistic awareness, concept formation, and creativity. These findings are based primarily on research with children in additive bilingual settings, that is, in settings where the second language is added as an enrichment to the native language and not at the expense of the native language. Causal relationships have been difficult to establish, but in general, positive outcomes have been noted, particularly in situations where bilingualism is not a socially stigmatized trait but rather a symbol of membership in a social elite.

Second-language acquisition

An important theoretical justification for the early view about the compensatory relationship between the two languages can be found in behaviorist accounts of language acquisition. If first-language acquisition consists of the establishment of stimulus-response connections between objects and words and the formation of generalizations made on the basis of the frequency patterns of words into sentences, then second-language acquisition must encounter interference from the old set of connections to the extent that they are different. The two languages were seen, in this empiricist account, as two sets of stimuli competing for a limited number of connections. This provided justification for the advice given to immigrant parents to try and use English at home so as not to confuse the children.

This empiricist account of language acquisition was strongly rejected in the late 1950s and 1960s on both theoretical (Chomsky, 1957) and empirical grounds (Brown & Bellugi, 1964). As with most revolutionary changes in the empirical disciplines, the nature of the questions about language acquisition changed in a qualitative manner. The new metaphor for the acquisition of language was the unfolding of innate capacities, and the goal of research became to delineate the exact nature of the unfolding process. If language acquisition was not the forging of connections between the stimuli of the outside world, then one would no longer have to see the learning of a second language as involving a "dog-eat-dog," competition with the first language. To borrow James Fallows's (1986) recent metaphor, having two languages is more like having two children than like having two wives.

There is considerable research support for this more recent view. For example, in the process of second-language acquisition, the native language does not interfere in any significant way with the development of the second language. Second-language acquisition and first-language acquisition are apparently guided by common principles across languages and are part of the human cognitive system (McLaughlin, 1987). From this structural point

of view, the learning of a second language is not hampered by the first. Furthermore, the rate of acquisition of a second language is highly related to the proficiency level in the native language, which suggests that the two capacities share and build upon a common underlying base rather than competing for limited resources (Cummins, 1984).

Language proficiency

Just as recent work in intelligence has moved away from regarding it as a single unitary construct (Sternberg, 1985), recent work on the notion of "language proficiency" has revealed a rich and multifaceted concept (Cummins, 1984; C. E. Snow, 1987). Research has extended the notion of language ability beyond grammatical skills to the use of language in various contexts, and more sophisticated notions are developing regarding language acquisition.

For example, C. E. Snow has identified at least two different dimensions of language proficiency in bilingual children. One dimension involves the use of language in face-to-face communicative settings (contextualized language skills), and the other dimension encompasses language use relatively removed from contextual support (decontextualized language skills). Contextualized and decontextualized language skills are independent, such that facility in interpersonal language use may not imply the ability to use the language in academic situations.

The diversification of language proficiency into different task domains complicates the task of understanding bilingual ability. The measurement of bilingualism has always been complex, and the maintenance of bilingualism in communities has been regarded by sociolinguists as best understood with respect to situational and functional constraints imposed in language use (Fishman, Cooper, & Ma, 1966). What is important is that language ability does not develop or atrophy across the board, that is, across the various domains of application.

Social context of language usage

Research on the use of the two languages in bilingual children (Zentella, 1981) suggests that they are adept at shifting from one language to the other depending on the conversational situation (a process known as code-switching) and that this behavior is not the result of the confusion of the two languages. Rather, bilinguals code-switch with each other to take advantage of the richness of the communicative situation, and from the viewpoint of ethnographers, one function of such code alternation is to establish and regulate the social boundaries of the two worlds (Gumperz, 1982). Such studies are important because they remind the student of child language that bilingualism (and language use in general) is a social phenomenon that takes place

between two or more parties and that questions of language use are really questions about social context, not about linguistic structure.

Conclusions about bilingual children

The research evidence suggests that second-language acquisition involves a process that, rather than interacting structurally with the first language, builds upon an underlying base common to both languages. There does not appear to be competition over mental resources by the two languages, and there are even possible cognitive advantages to bilingualism. It is evident that the duality of the languages per se does not hamper the overall language proficiency or cognitive development of bilingual children. Despite such conclusions, it is interesting to note the extent to which the debate over bilingual education has centered on the metaphor of languages in competition.

Bilingual education

The policy debate over how best to educate students who enter school with limited ability in English has focused on the issue of native-language support in instruction (August & Garcia, 1988; Baker & de Kanter, 1983). There is hardly any dispute over the ultimate goal of the programs—to "mainstream" students in monolingual English classrooms with maximal efficiency. The tension has centered on the specific instructional role of the native language: How long, how much, and how intensely should it be used?

On one side of this debate are supporters of native-language instruction. Proponents of bilingual education recommend aggressive development of the native language prior to the introduction of English. This approach is based on the argument that competencies in the native language, particularly as they relate to decontextualized language skills, provide important cognitive foundations for second-language acquisition and academic learning in general. The ease of transfer of skills acquired in the native language to English is an important component of this argument.

On the other side of the debate, some recommend the introduction of the English curriculum from the very beginning of the student's schooling experience, with minimal use of the native language. This strategy calls for the use of simplified English to facilitate comprehension. The approach is typically combined with an English as a Second Language (ESL) component. One intuitive appeal of this English-only method is its consistency with time-on-task arguments—that spending more time being exposed to English should aid students in their acquisition of English (Rossel & Ross, 1986).

Research and evaluation of bilingual education

Bilingual education programs have been in existence for over two decades, and thus the reasonable question arises as to whether there is evidence of the

relative effectiveness of the different approaches. Summative evaluations of programs that compare these different approaches have run into difficulty on a number of fronts. Willig (1985) in a meta-analysis of studies of the effectiveness of bilingual education, complained that evaluation research in this area is plagued with problems ranging from poor design to bad measurement. She concluded that "most research conclusions regarding the effectiveness of bilingual education reflect weaknesses of the research itself rather than effects of the actual programs" (p. 297).

The range of variability among the research approaches chosen is instructive. Almost all of the program evaluation studies concentrate on the effectiveness of the programs in teaching the students English, rather than focusing on students' overall academic development or factors other than traditional measures of school success Furthermore, the studies tend to observe children over only a limited duration, often no more than two years. The research defines its treatments and outcomes in strictly linguistic terms. At stake is the question of which approach would lead to faster and stronger acquisition of English. This question is a scientifically legitimate one, but it is dwarfed when compared to the outcomes that are of real long-term interest to society: the social and economic advancement of linguistic-minority populations through education.

Paulston (1980) expressed concern with the narrowness of the definition of program success in the following way:

> It makes a lot more sense to look at employment figures upon leaving school, figures on drug addiction and alcoholism, suicide rates, and personality discorders, i.e., indicators which measure the social pathology which accompanies social injustice, rather than in terms of language skills. . . . The dropout rate for American Indians in Chicago public schools is 95 percent; in the bilingual-bicultural Little Big Horn High School in Chicago the dropout rate in 1976 was 11 percent, and I find that figure a much more meaningful indicator for evaluation of the bilingual program than any psychometric assessment of students' language skills. (p. 41)

It is not always the case that English-language proficiency has guided educational research with bilinguals. The Significant Bilingual Instructional Features Study, funded in 1980, was a federal study that described instructional strategies in selected "effective" bilingual education classrooms around the country (Tikunoff, 1983). It was able to identify instructional attributes in these classrooms that were similar to those reported in effective nonbilingual classrooms as well as a set of attributes specifically common to the effective bilingual classrooms. More recent research, particularly that of Carter and Chatfield (1986) and Krashen and Biber (1988), has followed this earlier example of describing the organizational and instructional attributes of

schools and classrooms that produce academically successful bilingual students. However, even the most recent federal initiatives regarding program evaluation continue to look almost exclusively at English-language skills as the primary outcome variable (Ramirez, 1986).

Bilingual education policy

Continued focus on instructional language as treatment and English language as outcome can be directly traced to the judicial and legislative impetus for the development of programs and the related student eligibility criteria. The courts and Congress have repeatedly spoken directly to the disadvantages that students face as the result of their limited English proficiency. In the landmark 1974 United States Supreme Court decision in *Lau v. Nichols*, the court directly addressed the issue of language: "There a no equality of treatment merely by providing students with the same facilities, textbooks, teachers, and curriculum: for students who do not understand English are effectively foreclosed from meaningful education" (p. 26). In that same year, Congress addressed the issue in the Equal Education Opportunity Act (EEOA, 1974). The EEOA was an effort by Congress to specifically define what constitutes a denial of equal educational opportunity, including "the failure by an educational agency to take appropriate action to overcome language barriers that impede equal participation by students in its instructional programs" (EEOA, 1974, p. 1146).

Federal program initiatives in the form of targeted bilingual education legislation (in 1968, 1974, 1978, 1984, and 1988) have provided over a billion dollars in support for local school-district programs. In concert with the aims of the legislature and the courts, the main goal of these programs is to increase English-language proficiency. Guidelines for student inclusion in these programs have required evidence of limited English oral ability as assessed by a standardized English measure; a similar assessment of English proficiency is required prior to program exit. States with large numbers of bilingual students have adopted similar requirements. Moreover, these state and federal programs have focused their attention on the instructional strategies, frequently defined with respect to language of instruction, that will ensure the development of English-language proficiency.

The narrow linguistic definition of bilingualism in such programs has meant problems in accounting for all of the data. For example, as Cummins (1986) pointed out, linguistic mismatch between home and school may be a viable explanation for the school failure of some Spanish-speaking groups, but it fails to explain why some Asian-language groups have not experienced similar degrees of difficulty. Larger social and cultural factors embedded in the histories of different linguistic-minority groups may need to be taken into account (Ogbu & Matute-Bianchi, 1986), as well as differences in

learning styles that interact with instructional approaches (Wong Fillmore & McLaughlin, 1986).

That the linguistic definition of bilingualism in these programs can lead to imperfect predictability with respect to different groups of students should come as no surprise. Obviously, no quick fix for larger issues of social and cultural adjustment is likely to result from the manipulation of a single variable such as instructional language. We do not mean to suggest that the language variable is unimportant; rather, we are warning that the isolation of this single attribute as the only variable of significance ignores our present understanding of language as a complex interaction of linguistic, psychological, and social domains. The linguistic handle may have served policymakers well in focusing on an educationally vulnerable population of students, but it is clearly inadequate as the single focus of educational intervention aimed at ensuring academic competence for this population.

Future research

A considerable amount of knowledge has accumulated on bilingualism in recent years (summaries have been offered by Garcia, 1983; Grosjean, 1982; Hakuta, 1986; Haugen, 1973), and the topic has captured the attention of scholars from diverse disciplines. Inevitably, this body of research has overlapped with issues in education, particularly linguistic-minority education. The potpourri of concerns closely related to bilingualism constitutes a fertile meeting ground for social scientists with widely different research interests. We believe that future research should be directed at expanding the knowledge to be gained at the junctures of those diverse interests, as described in the following sections.

The language-cognition-affect connection

How language is related to general cognition and how both of these are involved with affective variables such as attitude, self-awareness, and identity formation can be fruitfully studied in bilingual individuals. Bilinguals, for example, provide test cases that disassociate variables in cognitive and language development that are otherwise conflated (Slobin, 1973). On the affective dimension, the relationship between affective variables and changes in language proficiency (e.g., greater degrees of acquisition of a second language or attrition of the native language) has been well explored in some settings (Gardner, 1983; R. D. Lambert & Freed, 1982). However, specific mechanisms about the relationship (e.g., Clark & Fiske, 1982) have yet to be proposed, and a coherent framework that takes into account issues of social identification processes (Gumperz, 1982) and emotions (Ervin-Tripp, 1987) must be developed. Bilinguals, as individuals who possess different

configurations of affect toward the two languages, provide important empirical evidence on such relationships.

Individuallsocietal levels of analysis

Bilingualism also offers an important area where the connections between individual and societal levels of a phenomenon can be studied. One example would be the notion of language vitality (Giles & Johnson, 1981) in individuals and in social groups. It is well known that bilingualism in social groups undergoes shift, often resulting in a monolingual community within two or three generations (Veltman, 1988). The rate of this language shift is a function of language vitality.

One argument for advocating aggressive development of the native language of linguistic-minority youngsters prior to introduction of English is that there is little environmental support for the home language because the social milieu (aside from the home and the immediate community) is overwhelmingly English (W. E. Lambert, 1984). Lower levels of language vitality at the larger community level presumably lead to lower levels of individual development in language proficiency. This relationship between the social milieu and the individual child has not been rigorously studied, but it provides an ideal "preparation" in which the impact of a societal level variable on individual development can be mapped out in detail.

Research, practice, and policy interface

There continues to be a great need for quality research on the basic processes of bilingualism as well as on the nature and effectiveness of educational programs that serve linguistic-minority students. The need is made greater because this topic readily invites "folk" speculation based, for example, on the experiences of immigrant relatives.

Among the various dilemmas confronting socially minded researchers is balancing responsiveness to the pressing need of society against standard scholarly attitudes toward applied research. Scientists with a sense of social responsibility often have resorted to bifurcating their energy, and scholars who have ventured into social policy have at times endangered their own scientific credibility. As in many areas of child development, bilingualism and education is an exciting arena in which base research can be conducted with educational and policy emphases, and with mutual enrichment rather than compromise (Zigler & Finn-Stevenson, 1987).

Indeed, in our view, scholars who conduct such research must step away from their traditional relationships with educators and policymakers. Rather than interpreting ivory tower research for practitioners, a collaborative structure and program of research must be formed through an ongoing dialogue between all parties involved in the education of linguistic-minority students, and new research questions can be generated from such discourse. An important

by-product of such collaboration would be the efficient translation of research into practical and political deliberations, as well as deep inquiry into the rote relationships between the various parties involved (Cummins, 1987).

Linguistic minorities and the linguistic majority

We believe that work in the area of bilingualism must establish continuities between the phenomenon as it occurs in minority and majority populations. For example is second-language acquisition in principle the same process when operative in linguistic minority and majority individuals? How is the acquisition of English by a Hmong refugee child different from the acquisition of French by a native speaker of English?

At the programmatic level, it is important to recognize the paradox that the educational system continues to convert linguistic-minority bilingual children into English monolinguals yet, at the same time, deplores the lack of competence of Americans in foreign languages many of which were natively spoken by minority children (Simon, 1980). So-called bilingual immersion programs (M. A. Snow, 1986), which combine language programs designed for minority students with those for majority students, should be encouraged and rigorously researched because they provide important continuity between the two groups and address an important societal need for a bilingually competent workforce.

Acknowledgment of the complexity

As we have argued throughout this article, the linguistic aspects of bilingualism provide only a window into a complex set of psychological and social processes in the development of bilingual children. A broad multidisciplinary perspective must be applied to the increasingly important problems faced by linguistic-minority students throughout the socialization process. How else are we to capture, understand, and respond to the sentiments of many immigrants, so eloquently expressed by this 10th-grade Chinese-born girl who had immigrated at age 12?

> I don't know who I am. Am I the good Chinese daughter? Am I an American teenager? I always feel I am letting my parents down when I am with my friends because I act so American, that I also feel that I will never really be an American. I never feel really comfortable with myself anymore. (Olsen, 1988, p. 30)

There is, indeed, more to issues confronting the bilingual individual than can be summarized by language proficiency measurements. As social scientists and educators, it is our obligation to capture the complexity of the situation and in the process to enrich our own science and practice.

Acknowledgments

We would like to thank Barbara Rogoff and two anonymous reviewers for helpful comments on an earlier draft of this article.

References

August, D., & Garcia, E. E. (1988). *Language minority education in the United States: Research, policy and practice*. Springfield, IL: Charles C. Thomas.

Ayres, L. P. (1909). *Laggards in our schools*. New York: Russell Sage Foundation.

Baker, K., & de Kanter, A. (Eds.). (1983). *Bilingual education: A reappraisal of federal policy*. Lexington, MA: Lexington Books.

Brown, R., & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. *Harvard Educational Review*, *34*, 133–151.

Carter, T. P., & Chatfield, M. L. (1986). Effective bilingual schools: Implications for policy and practice. *American Journal of Education*, *95*, 200–234.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Cark, M. S., & Fiske, S. T. (Eds.). (1982). *Affect and cognition: The Seventeenth Annual Carnegie Symposium on Cognition*. Hillsdale, NJ: Erlbaum.

Cummins, J. (1984). *Bilingualism and special education*. San Diego, CA: College Hill Press.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review*, *56*, 18–36.

Cummins, J. (1987). *Empowering minority students*. Unpublished book manuscript, Ontario Institute for Studies in Education.

Diaz, R. M. (1983). Thought and two languages: The impact of bilingualism on cognitive development. *Review of Research in Education*, *10*, 23–54.

Equal Education Opportunity Act of 1974, 42 U.S.C. §6705 (1975).

Ervin-Tripp, S. (1987, February). *La emoción en el bilinguismo*. Paper presented at the International Symposium on Bilingualism, San Juan, Puerto Rico.

Fallows, J. (1986, November 24). Viva bilingualism. *The New Republic*, pp. 18–19.

Fishman, J. A., Cooper, R. L., & Ma, R. (1966). *Bilingualism in the barrio*. Bloomington: Indiana University Press.

Garcia, E. E. (1983). *Early childhood bilingualism*. Albuquerque: University of New Mexico Press.

Gardner, R. C. (1983). Learning another language: A true social psychological experiment. *Journal of Language and Social Psychology*, *2*, 219–239.

Giles, H., & Johnson, P. (1981). The role of language in ethnic group relations. In J. C. Turner & H. Giles (Eds.), *Intergroup behavior* (pp. 199–241). Oxford, England: Blackwell.

Grosjean, F. (1982). *Life with two languages*. Cambridge, MA: Harvard University Press.

Gumperz, J. (1982). *Discourse strategies*. New York: Cambridge University Press.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism*. New York: Basic Books.

Haugen, E. (1973). Bilingualism, language contact and immigrant languages in the United States: A research report 1956–1970. In T. Sebeok (Ed.), *Current trends in linguistics* (Vol. 10, pp. 505–591). The Hague: Mouton.

Krashen, S., & Biber, D. (1988). *On course: Bilingual education success in California*. Sacramento: California Association for Bilingual Education.

Lambert, R. D., & Freed, B. F. (Eds.). (1982). *The loss of language skills*. Rowley, MA: Newbury House.

Lambert, W. E. (1984). An over view of issues in immersion education. In *Studies on immersion education* (pp. 8–30). Sacramento: California State Department of Education.

Lau v. Nichols, 414 U.S. 563 (1974).

McLaughlin, B. (1987). *Theories of second-language learning*. London: Arnold.

Ogbu, J., & Matute-Bianchi, M. E. (1986). Understanding sociocultural factors: Knowledge, identity, and school adjustment. In California State Department of Education (Ed.), *Beyond language: Social and cultural factors in schooling language minority students* (pp. 73–142). Los Angeles: Evaluation, Dissemination and Assessment Center, California State University.

Olsen, L. (1988). *Crossing the schoolhouse border: Immigrant students and the California Public Schools*. San Francisco: California Tomorrow (Fort Mason, Building B, San Francisco, CA 94123).

Paulston, C. B. (1980). *Bilingual education: Theories and issues*. Rowley, MA: Newbury House.

Ramirez, J. D. (1986). Comparing structural English immersion and bilingual education: First year results of a national study. *American Journal of Education*, *95*, 122–148.

Rossel, C., & Ross, J. M. (1986). *The social science evidence on bilingual education*. Boston: Boston University Press.

Simon, P. (1980). *The tongue-tied American: Confronting the foreign language crisis*. New York: Continuum.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York: Holt, Rinehart & Winston.

Snow, C. E. (1987). Beyond conversation: Second language learners' acquisition of description and explanation. In J. P. Lantolf & A. Labarca (Eds.), *Research in second language learning: Focus on the classroom* (pp. 3–16). Norwood, NJ: Ablex.

Snow, M. A. (1986). *Innovative second language education: Bilingual immersion programs* (Education Report 1). Los Angeles: Center for Language Education and Research, University of California.

Sternberg, R. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Thompson, G. G. (1952). *Child psychology*. Boston: Houghton Mifflin.

Tikunoff, W. J. (1983). *Significant Bilingual Instructional Features Study*. San Francisco: Far West Laboratory.

Veltman, C. (1988). *The future of the Spanish language in the United States*. New York: Hispanic Policy Development Project.

Willig, A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, *55*, 269–317.

Wong Fillmore, L., & McLaughlin, B. (1986). *Oral language learning in bilingual classrooms: The role of cultural factors in language acquisition*. Unpublished manuscript, School of Education, University of California, Berkeley.

Zentella, A. C. (1981). Language variety among Puerto Ricans. In C. A. Ferguson & S. B. Heath (Eds.), *Language in the USA* (pp. 218–238). New York: Cambridge University Press.

Zigler, E., & Finn-Stevenson, M. (1987). *Children: Development and social issues*. Lexington, MA: D. C. Heath.

82

CHALLENGING ESTABLISHED VIEWS ON SOCIAL ISSUES

The power and limitations of research

W. E. Lambert

Research on social issues often becomes mission oriented and focused on change, but rather than promoting immediate practical applications, it more often is limited to promises, new methodologies, and (sometimes) excitement for coresearchers within a discipline. In rare instances, however, social research does generate real power that contributes to palpable change. With the aim of clarifying the limitations and the power of such research, this article reviews a program of research, starting in the 1950s in Quebec, that attempts to improve interpersonal and intergroup relations in this ethnolinguistically diversified society.

According to D. N. Langenberg (1991), controversial public matters are shaped and controlled by an elite of civic leaders who are not scientists and who rely mainly on slogans rather than scientific knowledge. Consequently, societal realities are ultimately determined by societal perceptions, grounded in fact or not. Scientists shy away from social issues because societal perceptions and values are too easily transformed into societal realities, whereas "in science, *fact* yields perceptions, not the other way around" (Langenberg, p. 361). Yet Langenberg believes that it is an unfulfilled civic duty for scientists to become involved.

This call to civic duty grabs many of us who had experiences in the Great Depression and World War II and who got into the "softer" branches of science precisely because we wanted to improve on society by making it more fair. We believed we could have an impact through behavioral research if enough of us concerted our efforts; in time we would amass sufficient behavioral facts to redirect societal perceptions.

Source: *American Psychologist*, 1992, 47(4), 533–542.

This article is a retrospective assessment of the ups and downs of a program of research that was oriented by the ideals of societal change and fairness. The research involves me and numerous waves of talented undergraduate and graduate students and colleagues at McGill in Montreal from the 1950s on. In most cases, the clear limits of the social usefulness and value of the research we got into was apparent to all of us, but in certain instances we have been surprised and excited, and then anxious, because the research, whether worthy of it or not, had been believed and acted upon by local policy-makers, then across a nation and beyond. Perhaps a review of how one research group tried to challenge conventional slogans and wisdom will be informative enough to improve the ratio of hits to near hits or misses in social research. The article covers a range of interrelated issues, starting with studies of societal perceptions and an attempt to change stereotypes in multicultural Canada, and moving to the psychology and the social psychology of bilingualism, the role of attitudes and motivation in learning a second language, an attempt to engineer bilingualism in public schools, an empirical test of the putative distinctiveness of one's own cultural group, a comparison of enriching and debilitating forms of bilingualism, and finally an examination of how ethnic minorities and mainstream groups in North America cope with linguistic, cultural, and racial diversity in their communities. In each case there were many entrenched, conventional ideas to challenge, but the question of interest here is whether our challenges had a social impact.

An attempt to change stereotypes

It was not easy in Quebec in the 1950s to fathom the nature of the relationship between Canada's two "founding peoples," the francophones and the anglophones. The two groups were distinguishable mainly in terms of the social status positions held by members of each group and in terms of language usage. "One-way bilingualism" was the norm: The lower status francophones, composing some 80% of the provincial population, used English with varying degrees of skill, and they usually worked for anglophone employers who had no need, or urge, to learn French. The widely accepted description of intergroup relations was one of "two solitudes," after a moving and influential novel by a McGill professor, Hugh MacLennan (1945). The two-solitudes notion popularized a long-term social force in Canada that Hubert Guindon (1968) described as an accommodation by ethnic groups to the concept of ethnic cohabitation based on "mutually desired self-segregated institutions" that had in Quebec become "total in the fields of education, religion, welfare, leisure, and residence" (p. 707).

But solitudes meant more than separate worlds of experience; mutual ignorance of the other group was also implied, an ignorance sustained by powerful historical stereotypes that still have currency. Alexis de Tocqueville recorded in his notebook for August 29, 1831, that he had arrived in Quebec

at the precise moment of a "crisis—a problem of identity and a conflict about the assignment of political powers." Tocqueville marveled that the people in the countryside were "still French, trait for trait. They are like us, quick, alert, and intelligent." Yet he worried that "their natural charm put them in an inferiority position" before the anglos who, he found, were

> less likable but more efficient, cold, tenacious, pitiless, [whereas the] French seemed to enjoy what they have and often boast about what they don't have, more inclined to leave their families in need where they are themselves, trying to make out as well as they can.

In brief, Tocqueville painted a haunting portrait of the losers in early Canada. Eight years later, in a 1839 report of Lord Durham, an anglophone, who organized the union of the two Canadas, the plan of the English colonizers, the winners, is clear:

> Accustomed to hold a high opinion of their superiority, the English do not try to hide their distrust and aversion for the French Canadian's customs. They are leftovers of a former colony and will always be isolated in this Anglo-Saxon world. It is to pull them out of this inferiority that I want to give them our English character. However much they may fight against it, the process of assimilation to English customs is already underway. (Cited in Carron, 1977)

It still seemed to be underway in the Quebec of the 1950s, and reaction against it was becoming evident among the intellectual French Canadian elite, some of whom were vocally rejecting confederation and co-ethnic habitation and reconsidering a separatist alternative (see Guindon, 1968). Clearly, there were fixed positions about "the other group" that colored both facets of that double-solitude society. The temptation for social psychologists to probe these established views as extensively as possible was irresistible.

One incident helps set the stage. It involves a bus ride in the winter of 1958, when my attention was drawn to the conversation in front. One lady said something like, "If I couldn't speak English I certainly wouldn't shout about it," referring to the French conversation going on behind them. Her friend replied, "I know, but what can you expect?" They mentioned that they were especially bothered when French people laughed among themselves in their presence, as though they might be making fun of them—a case of linguistic paranoia. The whole flutter was prompted by a humorous conversation engaged in by two French Canadian women. The anglophone ladies couldn't understand the French conversation, nor did they look back to see what the people they seemed to know so much about even looked like.

From this and related experiences, we developed a research technique that makes use of language and dialect variations as a means of eliciting

stereotyped impressions. Briefly, listeners are asked to react to the taped recordings of a number of perfectly bilingual speakers reading a short passage at one time in one of their languages (e.g., French) and, later, a translation equivalent of the same passage in their second language (e.g., English). The listeners, kept unaware of the bilingual control, try to evaluate the personality characteristics of each speaker as well as possible, using voice cues only, much as we all do when we attempt to gauge the personalities of unfamiliar speakers heard over the phone or radio, sometimes in a language we don't know. This procedure, referred to as the *matched-guise* technique (Lambert, Hodgson, Gardner, & Fillenbaum, 1960), reveals judges' more private reactions to the contrasting group than do direct attitude questionnaires (see Lambert, Anisfeld, & Yeni-Komshian, 1965).

In 1958–1959, we asked English Canadian university students to evaluate the matched guises of male bilinguals speaking one time in Canadian-style French and a second time in English. As might be expected, their evaluations were strongly biased against the French Canadians and in favor of the matched English Canadians, whom they rated as being significantly better looking, taller, more intelligent, more dependable, kinder, more ambitious, and as having more character. This evaluational bias was just as apparent among judges who were bilingual as among those who were monolingual.

Fair enough, but when we presented the same set of taped voices to an equivalent group of French Canadian students, we were in for a surprise. They showed the same bias: They also judged the English Canadian guises significantly more favorably than the French Canadian guises on a longer list of traits. They saw the English Canadians as being taller, more intelligent, dependable, likable, and as having more character than the matched guises of members of their own ethnolinguistic group. On two traits only did they rate the French Canadian guises more favorably—kindness and religiousness—and in context it would seem that too much religion and kindness are questionable qualities. Not only did the French Canadian judges clearly down-grade French Canadian guises, they also rated the French Canadian guises much more negatively than did the English Canadian judges.

These results bothered us because they brought to light a deep-seated communitywide stereotype of French Canadians as being relatively second-rate people, a view apparently fully shared by certain subgroups of French Canadians themselves. Clearly, Tocqueville and Durham were not dead in the Quebec of the 1950s. In our write-ups we emphasized that the contrast drawn had no validity, that no individual could be more or less trustworthy, smart, dependable, or kind just because he or she switched languages, nor were French Canadians as a group actually shorter than English Canadians (they were not when we checked the heights of army recruits). But these caveats were not persuasive. For instance, several French Canadian colleagues interpreted the findings as proof that Anglo-Saxon cultural and linguistic domination had made French Canadian young people losers—an important

argument for separatist thinkers. Many English Canadian colleagues saw the results as proof of French Canadian inferiorities. Thus, rather than straightening out crooked thinking, or debunking conventional slogans about the members of the other group, this simple study seemed instead to justify stereotyped thinking because apparently everyone agreed about the personalities of English Canadians and French Canadians. Even the elites—the complacent English Canadians and the separatist-prone French Canadians—found solace in the results.

This research then missed its social target. But among researchers—the within-the-guild effect—it opened up exciting new leads. We could now check for consistencies through time and across social class, sex, and age groups (see Lambert, 1967). For instance, a follow-up study by Genesee and Holobow (1989) is very informative. Replicating the 1958 original study as closely as possible, their main finding was that, after a time lapse of some 30 years, there was no basic change in the essentially negative own-group views of French Canadian young people; they still rated French Canadian guises significantly less favorably than English Canadian guises on intelligence, dependability, ambition, education, and leadership ability. The durability of these sentiments is all the more noteworthy because apparently a "minority" group's social revolution for political and cultural independence could be initiated and sustained with little or no basic changes transpiring in own-group perceptions. Is it that there are political advantages to keeping pejorative own-group perceptions intact until full independence is assured? In terms of theories of self-esteem, there are few signs of self-enhancement in the French Canadian responses, no "self-protective mechanisms" (see Crocker & Major, 1989) that one might expect from a social minority group. Thus, questions arise about which group actually is the "minority group" in Quebec, and what it means to be a minority.

We had learned in this example that one doesn't change stereotypes by focusing research on them. This research brought us too close to raw sentiments. Another tack that would give us more perspective was called for, and we accordingly shifted our research attention to a more general question: How, in a purportedly multicultural nation like Canada, could one promote an open, multicultural vision of society, an appreciation of cultural and linguistic differences? In other words, we shifted our sights to multiculturalism and bilingualism, and in this case it gradually became apparent that research could effect conventional wisdom and established views and effect social change.

Exploring bilingualism

The first step was to explore bilingualism itself. In Canada in the 1950s not only was bilingualism "one-way" by nature, it was also broadly perceived as a plague associated with immigrant newcomers or with those on the margin of mainstream society. Internationally, the technical literature on the topic

strongly supported this perception: Numerous studies had apparently established that bilinguals, relative to monolinguals, were handicapped in terms of measured intelligence, school performance, and even social and personal adjustment (see Peal & Lambert, 1962). This amply documented negative overview became a slogan that could be used for various purposes. It could explain why French Canadians were the losers in Canada and why English Canadians were reluctant to become bilingual themselves. More generally in North America, the bilingualism–handicap message served the aims of assimilationists by frightening language minority groups and public school educators away from the bicultural–bilingual socialization of children.

When an odd study didn't find a bilingual handicap, authors became defensive. For example, when a study from the late 1920s found that Japanese and Chinese children from bilingual homes in Canada scored far above the White Anglo-Saxon norms on IQ tests, the authors cautioned that "the presence of so many clever, industrious, and frugal aliens, capable of competing (as far as their mental abilities are concerned) successfully with the native whites . . . constitutes a political and economic problem of the greatest importance" (Sandiford & Kerr, 1926).

Our research on bilingualism started out modestly. The size and scope of the technical literature had us convinced, but to fix or compensate for it, we needed to know what it is about being bilingual that constituted the handicap. Canada's federation of two founding nations had no place for assimilation through a single melting pot or single language. National unity depended on *bi*culturalism and *bi*lingualism. When we looked into the details of the more influential studies for clues about the causes of the handicaps (see Peal & Lambert, 1962), we found to our surprise that nearly all had commonsense inadequacies: The bilinguals were mainly from lower socioeconomic backgrounds or from rural areas, as in Wales; no checks were made on who was actually bilingual; and IQ and achievement testing was conducted in the monolingual children's language only. What amazed us was that this socially important issue was so insufficiently explored by so many established professionals from various nations over a 60-year period. Before searching for remedies for being bilingual, we now had to be certain that a handicap actually existed.

Our design, although not perfect, was an improvement, and most of the obvious flaws were corrected. The results were worth waiting for. Contrary to previous findings, bilingual 10-year-olds performed significantly better than did monolinguals (some 14 points on the average) on both verbal and nonverbal intelligence tests. All of a sudden the handicap was gone and replaced by a bilingual advantage: The bilinguals were more advanced in school, scored better on tests of first language skills, were more facile at concept formation, and displayed greater "mental flexibility" and a more diversified "structure" of mental abilities (Lambert, 1981; Peal & Lambert, 1962).

This study sent out a message that bilingualism was good. All that parents had to do was to make sure their children capitalized on opportunities to become fully bilingual. We even joked about it: If your child scores low on intelligence tests, make him or her bilingual! More seriously, we thought about the waves of immigrants to Canada, the United States, and elsewhere, who had been warned to protect themselves against bilingualism. In particular, we thought about one-way bilingualism in Canada.

We realized that this one study would be worrisome to many people; they would argue that the study was biased or that something peculiar was going on with Canadian youngsters. The best we could do was reassure ourselves on the main message (e.g., by testing different age groups and different language pairs). To have an impact, other researchers had to check us out in different cultural settings, and this happened rather quickly (e.g., Balkan, 1970, in Switzerland with French/English; Ben Zeev, 1972, in Israel and in the United States with Hebrew/English; Ianco-Worrall, 1972, in South Africa with African/English; Torrence, Gowan, Wu, & Alioti, 1970, in Singapore with Chinese/English). Each new study probed different aspects of what was already taken as a bilingual advantage, and surprisingly, no damaging criticisms have since appeared to disturb the basic change that had taken place in the conventional view of what it means to be bilingual. Here the power of social research became evident (see Lambert 1967, 1981). However, this new look of bilingualism, attractive as it may be in settings that favor a multicultural society, does not sit well with assimilationists or with separatists, which means that much follow-up research of a highly emotional sociopolitical nature can be anticipated.

What does it take to master a second or foreign language?

If being bilingual is good, Robert Gardner and I wondered how we might help young people, especially English Canadians, become bilingual. The technical literature on the development of second language skills focused mainly on individual differences in aptitude (e.g., that one has to have an "ear" for languages, or a musical ear, or an aptitude for math). Because little came from research checks on such ideas, interest shifted to verbal intelligence and related cognitive variables that were introduced into foreign language aptitude test batteries (see Carroll, 1973; Carroll & Sapon, 1959; Gardner & Lambert, 1959, 1972).

What interested us was that no one had examined carefully the social and emotional aspects of learning a second or foreign language. From our experiences in Quebec, we figured that success in mastering a foreign language would depend not only on intellectual capacity and language aptitude but also on the learners' perceptions of the other ethnolinguistic group involved, their attitudes toward representatives of that group, and their willingness to identify sufficiently to adopt distinctive aspects of behavior, like their

language. Learners' motivation and ultimate success in language study, we hypothesized, would be determined by attitudes and readiness to identify and by a positive orientation to the whole process of learning a foreign language. Thus, those with an "integrative" orientation—reflecting a sincere interest in another group of people and their culture—should learn better than those with an "instrumental" orientation—limited more to the practical payoffs of knowing a second language. Serious students intent on mastering the foreign language might confront a conflict of identity or alienation (we used the term *anomie*) as they became skilled enough to become accepted members of a new cultural group, calling for adjustments in allegiances (see Gardner, 1991; Gardner & Lambert, 1959, 1972).

These neglected aspects of language learning were tested out with English Canadian high schoolers studying French as a "foreign" language (Gardner & Lambert, 1959). Measures were taken of their language-learning aptitude and verbal intelligence as well as their attitudes toward French Canadians, their intensity of motivation, and their orientation to the learning of French. A factor analysis of these indices indicated that foreign language aptitude and verbal intelligence formed a single factor or cluster that—and this is the important point—was *independent* of a second factor made up of measures of motivation, type of orientation toward language learning, and social attitudes toward French Canadians. Scores on French achievement tests were reflected equally prominently in both factors. What this meant is that French achievement was dependent on a sympathetic orientation toward the other group as much as on verbal ability or aptitude. Moreover, the students' attitudinal orientations provided the motivation to learn the other group's language (i.e., those with an integrative orientation were more successful in learning French than were those who were more instrumentally oriented; see Gardner, 1991, for recent developments and Clément, 1978, for applications to French Canadians learning English).

At the societal level, however, there are worrisome features to this research. If attitudes and motivation have such an impact on second language achievement, the current attitudinal polarization of Canada's major ethnolinguistic groups makes the learning of the other group's language, and learning about the other group itself, that much less attractive or interesting (see Esman, 1987; Lambert, 1988). Canada's two solitudes seems to be changing into a case of double alienation, and that could turn both English Canadian and French Canadian youngsters off the learning of the other's language.

The main point, however, is that this research uncovered a new idea: Regardless of "natural aptitude," one can master a second language through an attitude–motivation route. The research thus challenged the conventional wisdom of those who stress the overriding importance of inborn aptitudes or God-given abilities. In this regard, it runs parallel to Benjamin Bloom's (1985) research on the development of talent in the fields of tennis, swimming, piano playing, mathematics, and so forth. World-class excellence in

these fields, Bloom argued, is apparently attributable less to native talent at the start than to tutors who capitalize on the interest, motivation, and attitudes of learners by helping, training, and then relaying the learners upward to more advanced tutors in order for talent to emerge and develop. This is a refreshing perspective that opens up exciting worlds of experience (from tennis to second languages) to all who *want* them enough, whatever their natural gifts. But much more research is needed before this important idea becomes a social reality.

Can one engineer bilingualism in a society of unilinguals?

If being bilingual isn't all that bad and if attitudes and motivation play appreciable roles in its development, how might one transform the one-way bilingualism and the mutual ignorance of the other group that colored Canada in the 1950s? The group most in need of becoming bilingual and knowledgeable about the other cultural group were the English Canadians, especially those who had decided to stay in the province of Quebec.

Our attempt at societal change in this case was through the education system, in which we, in collaboration with English Canadian parents, were able to initiate an innovative form of education, known as a "home-to-school language switch" program or as "immersion education" (Genesee, 1978, 1987; Lambert & Tucker, 1972; Swain, 1974). It was devised specifically for English Canadian children who were to spend most of the first three years of school almost exclusively in French, taught by French-speaking teachers. They were to be exposed to as much of the other language as possible, short of living in a French home. Our research role was to help in forming the program and then in monitoring its effects, especially the consequences of having little or no home-language instruction throughout the early elementary school years. Could these children become functionally bilingual and bicultural by the end of their elementary school years? Could they comfortably "add" a second language in this fashion?

The basics of immersion schooling are simple: English-speaking children, with little or no French language experience in their homes and communities, enter public school kindergartens conducted by monolingual French-speaking teachers. This language switch is kept exclusively French through the second grade, and only at the second or third grade is English language arts introduced for one period daily. Gradually, particular subject matters are taught in English (by a separate English-speaking teacher) so that by the fifth and sixth grades, some 50% of the instruction is conducted through English (see Lambert, 1979).

The underlying premise is that people learn a second (or third) language in the same way as they learn their first (i.e., in contexts in which they are exposed to it in its natural form and in which they are socially motivated to communicate. From the first encounter, immersion teachers use only the

target language. But from the start, the learning of language per se is made quite incidental to learning how to make and do new and interesting things. The new language becomes a constant verbal accompaniment rather than the focus.

Consistent findings from more than 25 years of research on these programs justify several conclusions: Immersion pupils are taken along by French monolingual teachers to a level of functional bilingualism that could not be duplicated in any other fashion, short of living and being schooled in a foreign setting; pupils arrive at that level of competence without detriment to their home language-skill development; without falling behind in the all-important content areas of the curriculum, indicating that the incidental acquisition of French does not hamper the learning of new and complex ideas; without any form of mental confusion or loss of normal cognitive growth; and without a loss of identity or appreciation for their own ethnicity. Most important of all, immersion pupils develop a deeper appreciation for French Canadians and feel closer and more similar to them by having learned about them and their culture through their teachers and through skills they develop with their language (see Genesee, 1987; Lambert, 1981; Lambert & Tucker, 1972; Swain, 1974).

Various side effects are fascinating. For instance, several studies (Barik & Swain, 1976; Cummins, 1975, 1976; Scott & Lambert, 1973) have found increases in students' IQs or in divergent thinking scores that are attributable to the development of bilingual skills through immersion schooling. Moreover, immersion education appears to enrich competence in English—the language used relatively little in the elementary grades. Recent studies (Lambert, Genesee, Holobow, & Chartrand, 1991; Swain & Lapkin, 1991), in fact, show significantly better performance for immersion students over matched control students in the upper elementary grades in various aspects of English skill and in content matters, such as math. Thus, it is clear that much across-language transfer takes place: Content learned through French percolates down to English; knowledge of the structure of French percolates down to the structure of English.

These outcomes contradict the previously conventional view that becoming bilingual—having two linguistic systems within one's brain—divides a person's cognitive resources and reduces efficiency of thought and language. Instead, there is now strong evidence for cognitive, educational, and social advantages to being bilingual (see Lambert, 1981, 1991) and also that these advantages are experienced as much by English Canadian children from working-class socioeconomic backgrounds as by the more advantaged, and for children with various levels of measured IQ, including children with diagnosed learning difficulties (see Genesee, 1987). Moreover, monolingual English Canadian children from Jewish families can handle easily and profitably a "double immersion" program, wherein two "foreign" languages are used for instruction from kindergarten through the elementary grades (see

Genesee & Lambert, 1983). The striking success of these double immersion programs in Montreal suggests that ethnic minorities in Canada might easily manage and enjoy education that is trilingual—French, English, and heritage language—just as the Jewish children in the double immersion programs not only manage but also enjoy education that is French, Hebrew, and English.

The immersion experience also fosters particular sociopolitical insights that monolingual mainstreamers would likely never develop—for example, that peaceful democratic coexistence in Canada calls for something more than simply "learning one another's languages." It also calls for opportunities wherein young people of both ethnic groups can interact in social, educational, and work settings on an equitable basis. Only then, these young bilinguals argue, can the mutual strangeness and suspicion be dispelled (see Blake, Lambert, & Sidoti, & Wolfe, 1981; Cziko, Lambert, Sidoti, & Tucker, 1980). This is a very sophisticated insight. Few of us monocultural adults who try to understand group tensions are able to see things so clearly.

This, then, is an example of a research hit in which researchers and a local group of enlightened parents were able to challenge conventional wisdom by initiating a radical educational experiment wherein children would be the catalysts for other-group tolerance and appreciation. It caught on and spread to other schools across the nation, to the United States (see Rhodes, 1991), and into Europe (see Lietti, 1989). In this example, the power of research makes many of us who were involved feel both proud and responsible. But even in this case, the tolerance and appreciation ran the risk of being one way, not mutual.

French Canada's response to two solitudes: experimenting with separation

Immersion education was an experiment shaped by English Canadians who saw possibilities for bilingualism and biculturalism at the local level and multiculturalism at the national level. French Canadians were not involved; they were debating the merits of political separation from the rest of Canada, with the focus on unilingualism and uniculturalism.

Noteworthy as the immersion program has been for anglophone Canadians, the social changes on the francophone side in the same time period are more remarkable, and in their case—and this is the main point—no psychologists, with detailed research plans, were involved. According to Milton Esman, a political scientist at Cornell, French Canadian intellectuals and professionals, from the late 1950s on, successfully "challenged the inevitability of francophone backwardness" and initiated "a new version of FC nationalism" known as the Quiet Revolution (Esman, 1987, p 397) In 1974, French was established as the sole official language of Quebec; in 1976, the separatist Parti Québecois was elected, which tightened the move toward

unilingualism for schooling and for professional licensing; French was instituted as the sole language of work in the province to the point that no store front signs in other than French have been permitted. In that time span, "the elites of the Quiet Revolution set in motion a remarkable modification in the ethnic division of labor. They have successfully employed their control of the state to increase the participation of francophones in middle-class managerial positions" (Esman, p. 401). The movement's success reminds all of us involved in research directed toward social change that, in the sociopolitical domain, there are very effective alternatives to the empirical type of research we psychologists rely on. The success of this momentous social movement also becomes a very heartening story for ethnolinguistic minorities elsewhere because it reveals clearly the deep attachments most people have to their heritage cultures and languages. But it also is an example of an attempt at social change that misses the target of mutual tolerance and appreciation, the hallmarks of constructive change in an open, ethnically pluralistic society. Instead, one sees a new French Canadian society that is passionate about its own culture and language, but one that is peculiarly without compassion for the other cultures and languages of its cocitizens.

Challenging the "distinctive" idea

Actually, there is nothing special about French Canada in this regard, if one considers the numerous independence movements currently underway around the world. Nor is there anything new about such movements, as similar campaigns for independence in the past have sliced the world up into an enormous checkerboard of nation states. Although little psychological research is available on the sentiments that underlie such movements, some features are obvious enough. When a subgroup within a nation, or some minority within that subgroup, talks about separation from the larger collectivity, it is probably a symptom that that group has been ignored or taken for granted, meaning the some basic needs for identity recognition have not been satisfied, creating a potent mixture of ethnocentrism and xenophobia.

Separatist movements are of concern to social psychologists, particularly when they promise a return to the good old days in one, big, happy family—symptoms that Karl Popper (1966) referred to as the "closing" of a society. Not only are ethnocentrism and xenophobia exaggerated but the belief that one's own group is distinctive is also made paramount, intensifying in-group versus outgroup thinking and the related belief that cultures and languages create distinctive personality styles and modes of thinking and unique ways of life.

A group of us in the 1970s wondered about the distinctive idea and attempted a simple test of it. Unfortunately, I offer it as another example of a

research miss. Granted that separatists believe and feel that they are distinctive, we wondered if this might be another instance of crooked thinking that we might straighten out. Briefly, we conducted a large-scale, cross-national study of parents' child-rearing values, assuming that distinctiveness should emerge on the basic issue of how different groups (such as French Canadians and English Canadians) bring up their children. Extensive interviews were conducted with separate samples of working-class and middle-class mothers and fathers of 6-year-old boys and girls in 11 national settings (Lambert, Hamers, & Frasure-Smith, 1979).

What emerged from this investigation, despite the variety of forms of parental values and attitudes toward child rearing in vogue in each of the nations studied, is the finding that the single most important influence on child-rearing values turned out to be the social-class background of parents, *not* their cultural or linguistic ethnicity. In fact, after social class, the sex of the parent and the sex of the child involved were both more powerful than cultural background. In Canada, for instance, the effects of social class were extensive, and they held as well for English Canadian as for French Canadian parents. Moreover, the values of French Canadian, English Canadian, and U.S. American parents formed a coherent North American regional pattern that makes each more North American in their value patterns than they are "American," "English Canadian," or "French Canadian." Thus, French Canadian and English Canadian parents of the same social-class background are more similar to one another in their child-rearing values than is either group to same-ethnic parents of a different social-class background. Similarly, British parents of a particular social-class standing are more similar in values to European French, Japanese, Greek, or French Canadian parents of the same social class than they are to British parents of a different social-class background.

What this research suggests to me is that this and possibly other forms of distinctiveness or uniqueness of one's own group (relative to other groups) may be more apparent than real. Between-group variance did not dominate within-group variance in this study of values. The main point, however, is that this research had very little influence. Few people were willing to believe the findings and their implications—even my friends and colleagues—probably because the findings contradict deeply engrained beliefs about own groups and out-groups that sustain their own ethnic identities. The best a researcher can do in this instance is to offer the study as a small strand of evidence and wait for an accumulation of strands that may in time become persuasive. For instance, the work of Steven Steinberg (1981) and an important study by Marvin Zuckerman (1990) add to such an accumulation. Zuckerman argued that "studies of temperament, basic personality traits, disorders (such as antisocial personality), and specific genetic markers show that there is much more variation within groups designated as races than between such groups" (abstract, p. 1297).

Education for ethnolinguistic minorities

Fortunately, there are happier, more open ways to satisfy the needs of language minority groups than the separatist way. We have referred to the process of developing the bilingual and bicultural skills of English Canadian children as an "additive" form of bilingualism, implying that these children, with no fear of ethnic or linguistic erosion, can add one or more foreign languages to their accumulating skills and profit immensely from the experience—cognitively, socially, educationally, and even economically (see Lambert, 1981; Lambert & Tucker, 1972), with no detriment to the home language or school performance. As we probed the features of additive bilingualism, we saw in sharp contrast its "subtractive" form, experienced by ethnolinguistic minority groups—the French Canadians included—who in many cases feel forced by social pressures or educational policies to put aside or subtract out their heritage languages for the more necessary and prestigious language of North America. Children in these cases can get caught in a psycholinguistic limbo, and we wondered how these debilitating aspects of bilingualism and biculturalism might be transformed into additive ones.

A few research-based examples of how this transformation might work have been tried out successfully (see Dubé & Herbert, 1975; Guzman, 1982; Lambert, Giles, & Albert, 1976; Lambert, Giles, & Picard, 1975). An exciting new alternative currently being tested in various sites integrates both mainstream and language minority children in "two-way bilingual immersion" classes, in which, for example, an anglophone and Hispanic teacher separately split the teaching day, with classrooms equally divided between English-speaking (Whites and ethnic minorities) and Hispanic youngsters. The early outcomes are extremely encouraging. In time this option will likely become a model because it offers "other language–culture" enrichment for mainstream Anglo children at the same time as it transforms the highly probable subtractive outlook of language minority children into an enriching and comfortable form of bilingualism and biculturalism (see Cazabon, Lambert, & Hall, 1991; Lindholm, 1990a, 1990b; Ramirez, Yuen, Ramey, & Pasta, 1991).

But much more research is called for because the *bi*lingual–*bi*cultural enrichment suggested by these recent studies is worrisome for those who think along a unilingual–unicultural line, that national unity requires melting-pot assimilation or, in the Quebec case, separation from contaminating neighbors. Even economically it can be upsetting, as Portes and Rumbaut (1990) suggested in their U.S.-based research:

> This resource (being bilingual) and its associated advantages can come to represent a serious threat to monolinguals who must compete in the same labor markets. It is for this reason that calls for subtractive acculturation—not English, but "English only"—find a receptive audience (in the U.S.A.). (p. 211)

Putting bilingualism, biculturalism, and multiculturalism together

My final research example is a simple pulse taking of the thinking now in vogue in North America on how best to accommodate the growing ethnic and linguistic diversity of its populace. The study emerges naturally from the whole cluster of previous investigations. It is also an example of research that is not likely to have an immediate social impact because it turns up unpopular, anxiety provoking findings that threaten conventional wisdom about the doctrine of assimilation.

Nonetheless, that doctrine merits examination. For example, fear that assimilation works too well drives French Canadians to experiment with separatism, whereas fear that assimilation doesn't work well enough drives Americans to extremes, such as "English-only" policies. Some nations develop important political parties that attempt to avoid the issue by trying to keep newcomers and minorities out (see Lambert, Moghaddam, Sorin, & Sorin, 1990). In fact, Dominique Moisi (1990) is concerned that Europe itself could, in the 21st century, "return to its past evil genius, fatally attracted by the dark temptations of xenophobia, racism, and jingoism, . . . using the exclusion of others to express its inner fears" (p. A23).

Has assimilation really worked in North America? It is difficult to see its success in the lives of established minorities, such as Blacks, Jews, and Asians. In fact, one gets the impression that America has actually evolved a policy of segregated Jacuzzis rather than one of a single integrated melting pot. The question then is, If more and more ethnolinguistic minorities find ways of being bilingual and bicultural, how comfortably will they fit in the North American scene?

The research approach Donald Taylor and I worked out was a direct one. We conducted a community-based investigation of the attitudes of parents in the United States toward ethnic diversity and intergroup relations. They were asked to consider the pros and cons of two contrasting ideological positions: *assimilation*, the belief that ethnic, racial, or cultural minorities should give up their heritage cultures and take on the "American" way of life, and *multiculturalism*, the view that these groups should maintain their heritage cultures as much as possible. They were to take a stand on this debate and also to express their attitudes toward the maintenance and use of heritage languages and toward relationships with other ethnolinguistic groups in the community. The study was conducted in a large American metropolitan area, diverse in cultural and racial composition. With the help of bilingual interviewers, we selected samples of parents who belonged to one or the other of the major ethnic groups living in the area: those with Polish, Arabic, Albanian, Mexican, or Puerto Rican backgrounds, along with samples of American Whites and Blacks. We concentrated on working-class samples in all cases and added a middleclass White American group for comparison purposes (Lambert & Taylor, 1990).

What surprised us about the outcomes was the degree of consensus and agreement we found within and among all but one of the key ethnic groups surveyed (the exception was the sample of working-class Whites): The rest strongly favored multiculturalism over assimilation as the best accommodation strategy for America. Even the Polish Americans, many of whom were third generation in the United States, solidly endorsed multiculturalism. The survey's clear message was that these samples of hyphenated Americans want opportunities for themselves and their children to juggle two cultures, that is, to become bicultural and bilingual Americans rather than giving up heritage cultures and languages in order to become "American." In short, they wanted members of their families to become "double breeds" rather than half breeds, comfortable with two cultures and languages rather than losing one in an exchange. As one Arab father put it: "There is no way my son could be my son if he were not Arab first. But we want him also to be as American as anyone else, starting with English and school work." Moreover, two established American groups, the working-class Blacks and the middle-class Whites, also supported a policy of heritage culture maintenance over assimilation. However, there were limits to their support: Extending heritage *language* use in public spheres, beyond the home and ethnic community, met with disapproval. Nonetheless, the working-class White sample was distinctive and alone in its rejection of multiculturalism and in its expression of negative, racist attitudes toward other ethnic and racial groups.

Our second major conclusion was that all parental samples, both ethnic minority and established American, see distinct advantages to being bilingual. The outcome was unanimous that being bilingual would mean not only better school performance for their children and a deeper sense of pride, but also more success in future careers in the world of work. Even the ethnocentric working-class White parents saw these advantages for their own children, although they did not believe ethnic minority children should become bilingual or bicultural.

The third major outcome was that most of the ethnic groups surveyed had kept the heritage language strong in their families, in some cases over a 30-year residence period in the United States. Moreover, three out of five of these groups had effectively passed the heritage language on to their children so that the latter were as native-like in the home language as their parents were and surpassed their parents in English language skills. Because the aim of all five ethnolinguistic groups was to provide full bilingual and bicultural competence for their children, in three cases the wishes of the parents were being satisfied. Only the Polish and Mexican parent groups showed the subtractive effects of assimilation in that their children were poorly skilled in the heritage languages. What is new, however, is that the other groups were able to help their children become comfortably bilingual and bicultural. Our guess is that this spirit can become contagious, that

through bicultural tolerance, multicultural tolerance on a nationwide scale can be generated (see Lambert, Mermigis, & Taylor, 1986).

Apparently, the time has passed when it can be assumed that assimilation is inevitable, or that cultural backgrounds gradually become "symbolic remnants" only (see Gans, 1979; Lambert & Taylor, 1990). A new alternative seems to be more attractive: Become as American as anyone else but don't lose your heritage identity in the shuffle—or for Anglo-Canadians, be as francophone and Québecois as you want but don't lose out on being Canadian in the process.

Note however that this research did not discover what may turn out to be a fundamentally new perspective on ethnolinguistic diversity in modern societies. It merely *uncovered* this new idea from ethnic groups who are courageously experimenting with a new creation, namely "don't lose a culture; go for two." Still it was heartening to notice that this grass-roots accommodation jibed nicely with what we in our own painstaking way had been learning about individual and societal forms of bilingualism and biculturalism.

In perspective

Where does the power of socially oriented research lie? Its limitations become all-too-frequently apparent, but even these give us clues about its power. Exacting and plodding as it must be, it permits us to challenge accepted truths, to ask upsetting questions, to probe in our own distinctive styles unchartered ways. It also permits us to challenge not only our research colleagues but ourselves as well, in order that we not kid ourselves in the conclusions we draw. Socially oriented research also teaches us that we must test the validity of each developing construct from a wide variety of different angles, forcing us out of purely psychological concerns to sociopolitical ones. Research hits that lead to observable social change are rare, and they seem to emerge when we have persevered sufficiently to have developed a network of interrelated constructs that appear to hold together. Ironically, when social changes come, they are likely to be motivated personally, not socially—people will try bilingualism for their children if the personal advantages for doing so become persuasive enough. In fact, they may even try multiculturalism and multifaceted tolerance if they become convinced that they need not lose anything in the process and that they might have much to gain.

Editor's note

Articles based on APA award addresses that appear in the *American Psychologist* are scholarly articles by distinguished contributors to the field. As such they are given special consideration in the *American Psychologist's* editorial selection process.

This article was originally presented as a Distinguished Scientific Award for the Applications of Psychology address at the 99th Annual Convention of the American Psychological Association in San Francisco in August 1991.

References

Balkan, L. (1970). *Les effects du bilinguisme francais–anglais sur les aptitudes intellectuelles* [The effects of French–English bilingualism on intellectual aptitudes]. Brussels, Belgium: Aimav.

Barik, H., & Swain, M. (1976). A longitudinal study of bilingual and cognitive development. *International Journal of Psychology*, *11*, 251–263.

Ben-Zeev, S. (1972). *The influence of bilingualism on cognitive development and cognitive strategy*. Unpublished doctoral dissertation, University of Chicago.

Blake, L., Lambert, W. E., Sidoti, N., & Wolfe, D. (1981). Students' views of intergroup tensions in Quebec: The effects of language immersion experience. *Canadian Journal of Behavioral Science*, *13*, 144–160.

Bloom, B. (1985). *Developing talent in young people*. New York: Ballantine Books.

Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *International Journal of Psycholinguistics*, *2*, 5–14.

Carroll, J., & Sapon, S. (1959). *Modern language aptitude test manual*. New York: Psychological Corporation.

Carron, A. M. (1977, February 24, 25, & 26). Une nation malade du Quebec [A nation fed up with Quebec]. *Le Monde*, pp. 1–2; p. 6; p. 3.

Cazabon, M., Lambert, W. E., & Hall, G. (1991). *An evaluation of the Amigos program*. Cambridge, MA: Cambridge Public School, Office of Bilingual/Bicultural Education.

Clément, R. (1978). *Motivational characteristics of francophones learning English*. Quebec City, Quebec, Canada: Presses de l'Université Laval.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*, 608–630.

Cummins, J. (1975). *Cognitive factors associated with intermediate levels of bilingual skills*. Unpublished manuscript, St. Patrick's College, Dublin, Educational Research Center.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, *9*, 1–43.

Cziko, G. A., Lambert, W. E., Sidoti, N., & Tucker, G. R. (1980). Graduates of early immersion: Retrospective views of grade 11 students and their parents. In R. N. St. Clair & H. Giles (Eds.), *The social and psychological contexts of language* (pp. 131–192). Hillsdale, NJ: Erlbaum.

Dubé, N. C., & Herbert, G. (1975). *Evaluation of the St. John Valley Title VII bilingual education program: 1970–1975*. Unpublished manuscript, Madawaska, ME.

Esman, M. J. (1987). Ethnic politics and economic power. *Comparative Politics*, *19*, 395–417.

Gans, H. (1979). Symbolic ethnicity. The future of ethnic groups and culture in America. *Ethnic and Racial Studies*, *2*, 1–20.

Gardner, R. C. (1991). Attitudes and motivation in second language learning. In A. Reynolds (Ed.), *Bilingualism, multiculturalism, and second language learning* (pp. 43–64). Hillsdale, NJ: Erlbaum.

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second-language acquisition. *Canadian Journal of Psychology*, *13*, 266–272.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Genesee, F. (1978). Scholastic effects of French immersion: An overview after ten years. *Interchange*, *9*, 20–29.

Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Cambridge, MA: Newbury House.

Genesee, F., & Holobow, N. (1989). Change and stability in intergroup perceptions. *Journal of Language and Social Psychology*, *8*, 17–38.

Genesee, F., & Lambert, W. E. (1983). Trilingual education for majority language children. *Child Development*, *54*, 105–114.

Guindon, H. (1968). Social unrest, social class, and Quebec's bureaucratic revolution. In B. R. Blishen, F. E. Jones, K. D. Naegele, & J. Porter (Eds.), *Canadian society: Sociological perspectives* (pp. 702–710). Toronto: Macmillan of Canada.

Guzman, G. (1982). *An exemplary approach to bilingual education*. San Diego, CA: San Diego Unified School District, ESEA Title VII Bilingual Demonstration Project.

Ianco-Worrall, A. D. (1972). Bilingualism and cognitive development. *Child Development*, *43*, 1390–1400.

Lambert, W. E. (1967). A social psychology of bilingualism. *The Journal of Social Issues*, *23*, 91–109.

Lambert, W. E. (1979). Language as a factor in intergroup relations. In H. Giles & R. N. St. Clair (Eds.), *Language and social psychology* (pp. 186–192). Oxford, England: Blackwell.

Lambert, W. E. (1981). Bilingualism and language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 9–22). New York: New York Academy of Sciences.

Lambert, W. E. (1988, May). *"Minority" language rights and education in Quebec*. Paper presented at the Cornell University symposium on Minority Language Rights and Minority Education, Ithaca, NY.

Lambert, W. E. (1991). And add your two cents' worth. In A. Reynolds (Ed.), *Bilingualism, multiculturalism, and second language learning*. Hillsdale, NJ: Erlbaum.

Lambert, W. E., Anisfeld, M., & Yeni-Komshian, G. (1965). Evaluational reactions of Jewish and Arab adolescents to dialect and language variations. *Journal of Personality and Social Psychology*, *2*, 84–90.

Lambert, W. E., Genesee, F., Holobow, N., & Chartrand, L. (1991). *Bilingual education for majority English-speaking children*. Montreal, Quebec, Canada: McGill University, Psychology Department.

Lambert, W. E., Giles, H., & Albert, G. J. (1976). Language attitudes in a rural community in northern Maine. *La Monda Linguo-Problemo*, *15*, 129–192.

Lambert, W. E., Giles, H., & Picard, O. (1975). Language attitudes in a French–American community. *International Journal of the Sociology of Language*, *4*, 127–152.

Lambert, W. E., Hamers, J., & Frasure-Smith, N. F. (1979). *Child-rearing values: A cross-national study*. New York: Praeger.

Lambert, W. E., Hodgson, R., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, *60*, 44–51.

Lambert, W. E., Mermigis, L., & Taylor, D. M. (1986). Greek Canadians' attitudes towards own group and other Canadian ethnic groups: A test of the multiculturalism hypothesis. *Canadian Journal of Behavioural Science*, *18*, 35–51.

Lambert, W. E., Moghaddam, F. M., Sorin, J., & Sorin, S. (1990). Assimilation vs. multiculturalism: Views from a community in France. *Sociological Forum*, *5*, 387–411.

Lambert, W. E., & Taylor, D. M. (1990). *Coping with cultural and racial diversity in urban America*. New York: Praeger.

Lambert, W. E., & Tucker, G. R. (1972). *Bilingual education of children: The St. Lambert experiment*. Rowley, MA: Newbury House.

Langenberg, D. N. (1991). Science, slogans, and civil duty. *Science*, 361–364.

Lietti, A. (1989). *Pour l'éducation bilingue: Guide de survie à l'usage des petits Européens* [For bilingual education: A survival guide for young Europeans]. Lausanne, Switzerland: Favre, S. A.

Lindholm, K. (1990a). *Academic and language achievement of Hispanic and Anglo students after three years in a two-way bilingual/immersion program*. Unpublished manuscript, San Jose State University.

Lindholm, K. (1990b, April). *Two-way bilingual/immersion education: Theory, conceptual issues, and pedagogical implications*. Paper presented at a symposium on Critical Perspectives on Bilingual Education Research, Phoenix, AZ.

MacLennan, H. (1945). *Two solitudes*. Montreal, Quebec, Canada: Duell, Sloan, & Pearce.

Moisi, D. (1990, May 29). A specter is haunting Europe: Its past. *New York Times*, p. A23.

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs*, 546.

Popper, K. (1966). *The open society and its enemies* (Vols. 1 & 2). London: Routledge & Kegan Paul.

Portes, A., & Rumbaut, R. G. (1990). *Immigrant America: A portrait*. Los Angeles: University of California Press.

Ramirez, J. D., Yuen, S. D., Ramey, D. R., & Pasta, D. J. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children*. Washington, DC: U.S. Department of Education.

Rhodes, N. (1991). *Total and partial immersion language programs in U.S. elementary schools*. Washington DC: Center for Applied Linguistics.

Sandiford, P., & Kerr, R. (1926). Intelligence of Chinese and Japanese children. *Journal of Educational Psychology*, *17*, 366–367.

Scott, S., & Lambert, W. E. (1973). *The relation of divergent thinking to bilingualism: Cause or effect*? Unpublished report, McGill University, Psychology Department.

Steinberg, S. (1981). *The ethnic myth*. New York: Atheneum.

Swain, M. (1974). French immersion programs across Canada. *The Canadian Modern Language Review*, *31*, 117–128.

Swain, M., & Lapkin, S. (1991). Additive bilingualism and French immersion education: The roles of language proficiency and literacy. In A. Reynolds (Ed.),

Bilingualism, multiculturalism, and second language learning (pp. 203–216). Hillsdale, NJ: Erlbaum.

Torrence, E. P., Gowan, J. C., Wu, J. M., & Alioti, N. C. (1970). Creative functioning of monolingual and bilingual children in Singapore. *Journal of Educational Psychology*, *61*, 72–75.

Zuckerman, M. (1990). Some dubious premises in research and theory on racial differences: Scientific, social, and ethical issues. *American Psychologist*, *45*, 1297–1303.

Part XVII

COMPUTERS AND MEDIA IN THE CLASSROOM

83

ANNOTATION: COMPUTERS FOR LEARNING

Psychological perspectives

P. Light

The research literature on the use of computers in support of learning is already vast. In this review, the focus is on children's learning, and on the way in which psychological theories of learning have informed (and, to a lesser extent, been informed by) developments in the field of computer-based learning. Associationist, constructivist, and social-constructivist approaches are explored, and issues of equity, access, and special learning needs are addressed. It is concluded that computers have led to, and will continue to lead to, significant changes in both what and how children learn.

Introduction

Interest in what computers might have to offer for learning goes back at least 30 years (e.g. Suppes, 1966), but the widespread availability of micro-computers in the 1980s provoked new interest amongst psychologists and educators in the issue of "computers for learning". Computers became not only *the* things to learn *about*, they were *the* things to learn *with*. Predictions were made to the effect that by the turn of the century computers would provide the major means of learning at all age levels and in all subject areas (Bork, 1980). Against such utopian enthusiasms lay a heavy weight of accumulated experience; many "new technologies" had been offered to education as panaceas in the past, and the judgment of many was reflected in the quip that "the only successful piece of educational technology is the school bus".

At the time of writing, some 15 years on, the situation has perhaps stabilised sufficiently to take stock. In the U.K., for example, most classrooms

Source: *Journal of Child Psychology and Psychiatry*, 1997, 38(5), 497–504.

and a high proportion of homes have reliable access to at least one computer. Most teachers have received some training in the use of computers, and the school curriculum envisages computer use in most subject areas (Crook, 1994). In this article, I shall not try to offer a review of the whole field of "computers for learning". Rather, I shall explore some of the ways in which psychological accounts of children's learning have contributed to, and been reflected in, the way computer-based learning has developed. I shall concentrate mainly, but not exclusively, on learning in the context of the school. Issues of access and equity raised by the advent of computers in the classroom and in the home will be discussed, and consideration will be given to what computers have to offer to children with particular learning needs.

Computers and instruction

The psychological tradition which has had the longest and strongest influence on the development of "computers for learning" is associationist learning theory. The hallmarks of this approach might be summarised as follows: (a) it focuses upon achieving some desired pattern of overt behaviour, (b) it sees the generation of desired behaviour patterns in terms of progressive "shaping" through small incremental steps, and (c) it sees the principal mechanism for achieving such shaping as being the reinforcement of correct responses through the reliable delivery of extrinsic rewards.

Learning theory reached perhaps its most influential expression in the operant conditioning research of Skinner and colleagues in the 1950s. Skinner took the view that human "teachers" left a great deal to be desired when it came to designing and carrying through systematic programmes of instruction, and saw in technology the potential for great improvement in educational method. Just as a well-designed "Skinner box" provided the controlled and predictable environment necessary for teaching pigeons or rats, so a "teaching machine" could be designed to carry out these same functions in respect of children's learning. Typical of his own early machines was one in which each frame presented a single unit of study material from which certain figures or letters were missing. The student had to supply the missing element using sliders on top of the machine. The student then turned a handle to bring on the next item, but if the previous answer was wrong, the crank would not turn.

Teaching machines became rapidly more complex and sophisticated. Most were mechanical devices, though some incorporated audio-visual presentation. Enthusiasm for teaching machines was premised on the conviction that if programs could be set up in advance by psychologists with expertise in learning theory, great improvements could be expected in the efficiency of children's learning. However, in practice there was nothing to stop anyone writing such programs, and most that were published were regarded by

Skinner (e.g. 1965) with disdain. Although studies comparing specific teaching machine applications with "conventional teaching" usually showed substantial advantages for machine instruction, they were subject to many methodological limitations (Holland, 1965). By the mid-1960s teaching machines had begun to fall out of favour, not least because of the limited scope and poor quality of many of the programs produced.

It is interesting to speculate whether the fate of this early flowering of educational technology would have been very different if microcomputer technology had been available at the time to support it. The large mainframe computers of the mid–1960s could only practicably have been used for teaching at school level if very substantial numbers of students had been sitting at terminals simultaneously, working through the same programs. Even apart from cost considerations, such regimentation of education was a less than attractive prospect. By the time cheap microcomputers began to appear on the classroom scene in the early 1980s, radical behaviourism had lost much of its force, both within and beyond psychology. Even so, some of the directions in which educational computing has developed show considerable continuity with associationist approaches to learning.

Many contemporary uses of the computer to support learning fall under the heading of "computer assisted instruction" (CAI). The idea is that the computer "assists" the teacher to achieve an instructional goal. Indeed, as far as is practicable the computer functions *as* a teacher. The classroom teacher may be dealing with 30 children, and cannot give the kind of dedicated and sustained attention to the individual learner that would be required to implement an individualised instructional program.

CAI offers the prospect of individualised instruction, geared to each learner's level and pace. CAI software typically places heavy reliance upon "drill and practice". A good deal of it is focused upon the inculcation of skills in elementary arithmetic, spelling, etc., though subject matter extends right through to, for example, statistics for psychology students. Typically, large numbers of items are presented in a fixed format with gradation of difficulty, and with progression dependent upon correct responses. "Reinforcement" typically takes the form of congratulatory messages, but today's graphics allow these to be supplemented by all manner of diverting "sideshows" (in one case, for example, the student's successes result in the nudging of a screen image of the teacher closer and closer to a vat of custard, into which she eventually falls!).

The flexibility and power of the microcomputer, as compared to the earlier teaching machines, facilitates the provision of more feedback in respect of incorrect responses. Skinner saw reinforcement of correct responses as the key to progress, and advocated ignoring incorrect responses. However, today's students are likely to receive corrective guidance according to what particular error they have made. Similarly, whereas Skinner's programs were essentially linear, consisting of one long sequence of items through which all

students would proceed at their own pace, most programs used today are highly "branched", with the response to a given item being used as a basis for selecting the following item.

This general approach dominated computer use in education in the early years, and its influence continues strongly today. Software in the CAI tradition is probably more widely used than anything else in contemporary schools, at least for children under about 9 years of age (Crook, 1994). How successful is the use of this type of computer instruction, compared to more traditional approaches? Individual studies typically involve a comparison between a particular piece of CAI software and a particular form of traditional instruction (typically "chalk and talk", often supported by a workbook). The arbitrariness associated with the choice of comparators, together with the restriction to particular populations and contexts, limit the value of such research. However, reviews and meta-analyses of the large number of such studies conducted in the 1980s suggest that CAI is at least moderately effective. Niemiec and Walberg (1987), for example, conclude that the average (50th percentile) CAI student reaches a level of achievement that would put them at the 66th percentile in traditionally taught groups.

In most CAI software, the level of "error diagnosis" actually undertaken by the computer is minimal. An obvious direction for development of tutorial software is towards programs that can take account of the *pattern* of errors made by an individual child over time, building a model of how the child is reasoning about the problem in hand. This model of the learner's current behaviour could then be used to select the most appropriate path for further instruction. Such an "intelligent tutoring system" (ITS) would thus be "learning about the learner" in order to teach. In practice, systems developed to date rely either on supposing that the user's knowledge is simply a subset of adult knowledge, or on analysing deviations from correct responses using a list of "buggy" or incorrect rules. The former is clearly inadequate, whereas the latter is practicable only in extremely circumscribed domains, such as elementary subtraction (Costa, 1991).

It is perhaps a poor reflection on the psychology of cognitive development that it has not been able to offer accounts of children's learning at a level of detail needed to serve the development of intelligent tutoring systems better. However, even in areas where adequate models of children's learning are available, it has yet to be demonstrated that an ITS is necessarily more effective in practice than an "unintelligent" tutoring program based on the same learning model (Nathan & Resnick, 1994).

This section has spanned a considerable range of approaches to the use of machines for supporting learning, from Skinner's teaching machines to ITS developments of recent years. What these approaches have in common is a focus on learning as the product of a process of instruction, with the machine cast in the role of instructor. In the next section I will turn to approaches which find their starting points in a very different view of learning.

Computers and construction

The decades that witnessed the crucial early developments in the technology of microcomputers also saw the peak of Piagetian influence in developmental psychology and the psychology of education. From this standpoint, learning is seen as secondary to, and dependent upon, progressive reorganisations of the child's cognitive functioning. The key to education lies not in "shaping behaviours", but in providing the conditions in which children themselves can construct an understanding of the world in which they are growing up. The role of the teacher becomes essentially facilitatory, providing the environment, the resources, and the encouragement necessary for the child's creative explorations.

The influence of Piagetian theory on the development of "computers for learning" was mediated principally through the work of Papert, who in the late 1970s began to see in the computer a possibility for radically subverting the traditional methods of education. Papert (1980, 1994) argues that the potential of the computer cannot be realised by casting it in the role of surrogate teacher, controlling children's learning. On the contrary, its potential lies in extending children's control over their own learning. The computer should not be programming the child, the child should be programming the computer. The design of a modular programming language called Logo, suitable for use by even quite young children, lay at the heart of Papert's enterprise.

The most widely developed uses of Logo have involved "turtle geometry". Here a wheeled "turtle" robot (or more frequently these days a screen icon) is controlled from a keyboard, its movements being governed by a program written by the child. The structure of the program itself provides a concrete instantiation of the underlying geometry. Papert sees the activity of navigating the turtle as making the geometry a part of the child's "lived experience". More generally, programming in Logo is seen as linking intuitive and abstract levels of understanding and thereby fostering "powerful ideas" in the form of generalisable thinking and problem-solving skills. Writing a program to make the computer do something depends upon arriving at a schematic representation of the desired activity. The program is not judged right or wrong according to the canons of adult authority; it either works or it doesn't. If it doesn't, a process of "de-bugging" is required, which involves critically re-examining the schematic representation and its implementation. In contrast to learning theory, the child's learning activities are envisaged as being entirely intrinsically motivated.

Logo has come to be very widely used in schools, but evaluations of its use lend little support to the radical claims made by for it by Papert. There is no systematic evidence that unstructured Logo use changes the way in which children approach problem solving in other settings (Pea & Kurland, 1984). Only when Logo learning is sustained over a substantial period

(50–60 hours, say) with active teacher input and a highly structured curriculum is there evidence of generalisable gains in problem-solving abilities (De Corte, Verschaffel, & Schrooten, 1992). On the other hand, "turtle geometry" does provide an effective way for children to learn some aspects of mathematics (Hoyles & Sutherland, 1989), and it is at this more prosaic level, for the most part, that Logo justifies its existence in today's classrooms. The idea that programming represents the key to appropriate use of computers for learning has undergone a number of revivals. For example, "logic programming" with Prolog enjoyed a brief flowering (Nichol, Briggs, & Dean, 1988). Today it is more likely that children will encounter high-level authoring packages that allow them to create complex multimedia programs with relative ease. These are powerful tools, but not tools designed, as Logo was, to foster "powerful ideas".

Schank and Cleary (1995) share with Papert an emphasis on "learning by doing", and a conviction that the computer is the key to extending, throughout the school years, the natural and spontaneous forms of learning children show in the preschool years. However, as advocates of case-based reasoning, they see generalisation as emerging from substantial experiences of particular instances, or cases, and this leads them to see *simulation* as being a key function for computers in learning. Just as flight simulators assist pilots to learn to fly by providing realistic environments without all the risks and costs of the real thing, so a great range of other experiences can be simulated in computer-supported environments, allowing children to play, experiment, and explore. In this way the range of things children can "learn by doing" is potentially greatly extended, with teachers acting as "mentors", asking questions, making suggestions, and so on.

To give one example, a relatively simple simulation designed for American school children allows them to take simulated trips around the United States by car. On arriving at each destination the user can select amongst video clips illustrating a range of aspects of that location. Children navigate using country, state, and city maps, and get a view of the countryside as they pass through it on their way to their destination. Using this simulation. Schank and Cleary (1995) report that scores on conventional geography tests improved significantly for weak students, and a little for stronger students. More importantly, as the authors see it, the students developed the ability to *use* road maps to plan journeys; something not even tested by the usual assessments. Such simulations are clearly highly motivating, and fit in with the central idea that learning should be fun.

Using video, audio, graphics, and animation, computer simulations can let children explore the inside of the human body, or even the inside of a cell. They can also simulate "impossible" worlds. Thus the child may be helped to appreciate Newton's laws of motion by being able to try moving around in outer space. Equally, though, a space can be simulated in which *different* laws of motion apply (Smith, 1986). The development of such computer

simulations is expensive and there seems little immediate prospect of their having major impact in education at school level. However, developments in the technology of "virtual reality" may change this situation over the next decade, in which case the emancipatory potential of this mode of learning may yet get a fair test.

Computer modelling software is related to simulations, but provides more or less "content free" resources that can allow children to build their own on-screen models of anything from the rotation of planets to the economics of the corner shop (diSessa, Hoyles, & Noss, 1991; Mellar, Bliss, Boohan, Ogborn, & Tompsett, 1994). In this way, children can be encouraged to render complex and half-understood phenomena into a concrete model, which can be "run" to see what happens and then edited accordingly.

Though less specifically linked to a constructivist agenda, the advent of Hypertext and Hypermedia environments still puts the emphasis upon "putting the learner in control" (Hutchins, Hall, & Colbourn, 1993). Such environments comprise rich learning resources relating to a given topic, in the form of "blocks" of information that may take the form of a piece of text, a picture, a video clip, or a sound sequence. The blocks of information can be accessed in any order by the user, who can move between them at will using "links". Links may be pre-programmed or user-defined (Yankelovich, Meyrowitz, & van Dam, 1991). At limit, there may be no linear structuring of the material at all, with everything left to the user's motivation to explore. In practice, varying degrees of "guidance" are usually provided within the software.

This section has considered a range of developments in the field of "computers for learning" that have in common a more or less specifically constructivist vision of development and learning. In most, the focus has been on the individual learner in interaction with the machine. However, just as Piagetian constructivism in developmental psychology has been overtaken by a "*social* constructivist" agenda often identified with Vygotsky, so in the field of educational technology the social dimensions of computer use have become a major focus of attention in recent years.

Computers and collaboration

As we have seen, advocates of the use of computers for learning have often argued in terms of the individualisation of learning. At the same time, critics have tended to see such individualisation of learning as threatening the social fabric of education. Each plugged into their own machine, children would have no need or opportunity for social interaction. In reality, things have worked out rather differently. Computing resources in schools have always tended to be scarce, and for this reason if no other, children have usually been set to work at the computer in pairs or small groups, often working relatively independently of the class teacher. Moreover, it appears that groups

focused around a computer task do rather better than most other sorts of groups in terms of sustaining task-related interaction over long periods (Scanlon, Murphy, & Issroff, in press).

Groupwork in the classroom has always had its supporters (Foot, Morgan, & Shute, 1990), some of them influenced by Piaget's argument that peer interaction offers a more potent stimulus to cognitive development than teacher instruction (Doise & Mugny, 1984). Amongst contemporary developmental psychologists, enthusiasm for collaborative modes of learning more often reflects the influence of Vygotskian ideas (Newman, Griffin, & Cole, 1989). While sharing many of Piaget's ideas about cognitive development as an active process of construction, Vygotsky envisaged such construction as an essentially social process, with *other people* representing a key resource for learning. Teaching and learning are conceived of largely in terms of informal processes of apprenticeship in which more expert individuals "scaffold" or support the performance of less expert individuals. Both the interpretation of a task and the development of a solution are envisaged as processes of collaborative co-construction, processes which may go on between collaborating learners just as they may between teachers and learners.

The advent of computers gave new impetus to research on collaborative learning. Most studies in the programmed learning/computer-assisted instruction tradition have indicated that pairwise instruction at least is no *less* effective than individual instruction (Elshout, 1992), and some have pointed to both cognitive and affective benefits in small group as compared to individual instruction (Mevarech, Silber, & Fine, 1991). Logo has been found to lend itself well to collaborative modes of learning, and such collaboration has been found to be of benefit both in relation to mathematics learning (Healy, Pozzi, & Hoyles, 1995) and in terms of cultivating higher-order thinking skills (Clements & Nastasi, 1992). Once again, though, there are other studies that have shown no such benefit from collaborative modes of working (Hughes & Greenhough, 1995).

Our own research with information searching and planning tasks couched as adventure games has found advantages for paired as against individual modes of learning both in terms of group performance and in terms of subsequent individual performance on similar tasks (Blaye, Light, Joiner, & Sheldon, 1991). However, the children in the "control" condition were working alone in a room on an unfamiliar computer. In further studies we found that, against this baseline, the presence of other children in the room was facilitative even if the children were working on separate computers without any interaction. Thus the effects may owe as much to affective and motivational factors as they do to any direct benefits of interaction (Light, Littleton, Messer, & Joiner, 1994).

Collaborative learning using computers can also take place at a distance. It may be mediated via computers, using networks. For example, computer simulations have been designed in which many users at many different sites

can participate simultaneously (Smith, 1991). Computer-supported collaborative learning is a growth area (O'Malley, 1995). Its relevance is perhaps most obvious in relation to distance learning (Mason & Kaye, 1989), but electronic mail and computer conferencing may turn out to have important implications for education even in the context of full-time educational settings (Crook, 1994; Light & Light, in press; Robinson, 1993).

Research on collaborative modes of computer use can be seen as a step along the path towards considering the wider social and institutional context within which learners are encountering computers. Psychological research on learning all too often neglects the roles of teachers and peers, and the wider social and institutional frameworks within which learning occurs. In the case of computer-based learning we have only recently begun to see the emergence of detailed qualitative research on computer use in authentic learning environments (Crook, 1994; Mercer, 1995, Schofield, 1995). This research arguably reflects the true legacy of Vygotsky's "sociocultural" account of development, reaching beyond a description of patterns of interpersonal interaction around the computer towards an understanding of the fuller learning context of which the computer forms a part.

Another aspect of a socio-cultural approach to learning with computers is attention to issues of differential access and opportunity. A technology which affords opportunities for learning may tend to sustain or exacerbate existing inequalities of access, but equally it may have the potential to ameliorate them. These issues will be addressed in the next section.

Equity, access, and special learning needs

One of the "downsides" of the use of computers for learning is that girls often seem to be a good deal less enthusiastic about their use than boys. U.K. surveys suggest that more girls than boys have negative attitudes towards computers, and that girls participate less than boys in various computer activities in school (Culley, 1993). Differences seem to become more pronounced the longer the children are in school. Also, as indexed by subject choice at secondary and university levels, it appears that this gender imbalance in response to computers has become stronger over the last decade or so (Newton & Beck, 1993). International comparisons show that gender imbalance in response to computers amongst students is widespread (Pelgrum & Plomp, 1993). Home computers are also more likely to be bought for boys than for girls, and where they are available, boys tend to use them more than girls, for all purposes (Martin, 1991).

There is little evidence that boys actually perform better on computer-based learning tasks than girls, and when they do it seems that gender stereotyping in the software is often to blame. For example, we have found in our own work that superficial modifications to the scenario used for a computer-based problem-solving task can remove gender differences in

achievement (Littleton, Light, Joiner, Messer, & Barnes, 1995). In mixed gender groups boys will tend to dominate the mouse or keyboard, but it is not clear how adversely this affects girls' learning. In a recent study we have found that boys and girls in mixed gender groups show similar learning outcomes, whereas the presence of a child of the opposite sex working alongside on a separate machine has positive effects for boys, negative for girls (Light, Littleton, Bale, Joiner, & Messer, in press). Social comparison effects thus appear to exert a powerful influence in this situation.

Gender differences in "style" have been noted in various studies of children's learning with computers. Girls are held to prefer collaborative modes of working, and adopt a more open-ended and exploratory approach, whereas boys prefer to work alone, and adopt a more analytic and closed approach (Turkle & Papert, 1990). Either style is compatible with using computers for learning, but it may be that, to date, software development for learning as well as for leisure has tended to follow a "masculine" path (Kirkup, 1992). Concerns about equity and access have also arisen in relation to race and socioeconomic status. Implicit assumptions about the capabilities of black and lower SES children may be reflected in the "drill and practice" type of software to which they are predominantly exposed (Scott, Cole, & Engel, 1992).

Children in many less-developed countries have no access to computers at all. In education, as in other fields, the information technology "revolution" has the potential to widen the gulf between rich and poor countries dramatically, as the industrial revolution did in an earlier era. If they could develop the necessary skills, of course, developing countries could benefit from the fact that computer and communications technologies make it possible to shift employment around the world much more easily. The education budgets of some developing countries, however, scarcely extend to buying children basic reading and writing materials, let alone computers (Hawkridge, Jaworski, & McMahon, 1990).

Another area in which economic factors may exert a major influence concerns the purchase and use of home computers. There appears to have been a good deal less research on the use of computers for learning in the home than in the school environment, and what there is has mostly been restricted to white, middle-class families. In one such study, Giacquinta, Bauer, and Levin (1993) found that almost half of their sample of 70 families had purchased the computer with their children's educational needs uppermost in their minds. However, their research revealed only a modest amount of general educational use (e.g. word-processing) and almost no use of the computer to support study of school subjects. The authors point to a lack of liaison between home and school and a lack of willingness on the part of parents (especially mothers) to get involved with the computer as being the main contributory factors.

Predictably, Giacquinta et al. (1993) observed that the home computers were used mainly for game playing. Computer games tend to be regarded with little enthusiasm by parents and teachers. However, some psychologists have attempted to rescue computer games from opprobrium by suggesting that they can foster a range of useful skills (Greenfield, 1984). Meanwhile, the evident success of computer games in motivating children has led to a number of deliberate attempts to use game formats to teach particular skills, and indeed much educational software has been influenced in some measure by computer games. The resulting growth of "edu-tainment" software may in the long term lead to a more successful integration of learning at school and at home than has so far been achieved. A danger in the short term is that the strong gender stereotyping associated with today's computer games could aggravate any existing disenchantment that girls may feel towards computers at school.

Children with special educational needs are a particular group for whom computers can offer real enhancements of access to learning. This is especially true for individuals suffering physical disabilities. Though we tend to think of computers in terms of keyboard, mouse, and screen, the range of possible devices for interfacing with computers is virtually limitless. Any bodily movement over which the individual has control, even a movement of an eyebrow, can in principle be used to control a microcomputer. A great deal of ingenuity has been put into developing appropriate interfaces for such purposes.

For students with visual handicaps, screen magnification software and "screen readers" that speak the text are available. Hearing-impaired children can benefit from computer use in relation to language learning (Gray, 1995). Electronic means of communication (including Email, conferencing etc.) have particular advantages for those with speech difficulties or severely limited mobility (Coombs, 1989). Voice recognition software is rapidly improving.

"Smart" wordprocessors, which anticipate the text to be entered on the basis of the first few letters and adapt to the lexicon of the user, can benefit physically disabled children by speeding up text production. Synthesised speech in the context of a multimedia approach to augmentation of written text seems to offer particular benefits to dyslexic children (Fawcett, Nicolson, & Morris, 1993).

Children with general learning disabilities form a large constituency, but despite some early advocacy (e.g. Goldenberg, 1979) computers have not as yet been widely used with these children (Hawkridge & Vincent, 1992). Initially, computer-assisted instruction of the "drill and practice" type was the most likely type of software to be used, and indeed there is evidence of its effectiveness, for example in helping to overcome reading problems (e.g. Torgesen & Barker, 1995). In recent years an increasing range of computer

applications has been put to use. For example, Logo has been used to encourage mathematical understanding even with children with severe learning difficulties (Hawkridge & Vincent, 1992). Computer-based dialogue involving teachers and learning-disabled students seems to have some potential to reduce the strongly didactic forms of discourse commonly adopted in the classroom with such children (Rueda, 1992).

Autistic children are reported to find working with computers particularly congenial. Swettenham (1996) attributes this to the asocial aspect of working with a machine, the consistency and predictability associated with the computer, and the fact that the children can work at their own pace. However, direct comparison of personal and computer-based instruction for such children suggests that, although computer use may have advantages in terms of motivation and behaviour, learning outcomes are not likely to be dramatically affected (e.g. Chen & Bernardopitz, 1993).

Use of computers with children suffering emotional and behavioural difficulties has typically focused upon improving the children's image of themselves as learners by enabling them to achieve success in what is a highly regarded area (Hopkins, 1991). A rather different approach focuses on the potential of the computer to foster self-expression and communication. For example, Bubble Dialogue (O'Neill & McMahon, 1991) encourages children to role play within a comic-strip environment, in an unthreatening "computer game" setting that provides some distance between them and the characters. Its use for stimulating and capturing children's "exploratory talk" appears to have potential, in both educational and therapeutic terms (Jones & McMahon, 1994; Jones & Selby, 1995).

Conclusions

As is evident from this review, computers have been of interest to psychologists of many persuasions. Innovations in adapting and exploiting computer technology in support of learning have often originated in the work of university research teams, and psychologists have, directly and indirectly, played a significant role in shaping many of them.

As we saw at the outset, the teaching machine tradition envisaged replacing inefficient classroom teaching with psychologically informed programs of individualised instruction. Work on computer-assisted instruction and intelligent tutoring systems still has traces of this stance. The idea that the computer will "transform education into psychology" (Sardello, 1984) has seemed to some a vision, to others a nightmare. However, in practice, the attempt to use computers as surrogate instructors has had only limited success, and has been largely superceded by attempts to use the computer to provide rich and stimulating learning environments for children.

Constructivist-inspired software potentially suffers the same problems as constructivist theories of learning more generally, namely an idealised notion

that intrinsic motivation will always be sufficient to promote learning, and a lack of attention to the social dynamics of learning situations. However, software developed through this approach (from Logo through simulations and modelling tools to Hypertext resources) has typically been assimilated rather pragmatically into the social and motivational practices of the classroom. Cheaper computer memory and improving access to broad bandwidth networking make it likely that the computer in the classroom will serve increasingly as an access route to information rather than as a substitute teacher.

For many people in developed societies, computers have come to function simply as tools for doing everyday tasks. Word processors, spreadsheets, databases, and graphics packages have become part and parcel of daily life. The use of computer-based tools of this kind has changed the nature of work in many walks of life, and in doing so it changes what children *need* to learn as well as changing the means available to learn it. The advent of the word processor, for example, subtly alters the requirements of literacy (Saloman, Perkins, & Globerson, 1991). In mathematics, computers alter what arithmetical skills are needed as well as providing different means for learning them. In music, software packages offer routes to both the composition and production of music that depend on few of the skills traditionally associated with learning music. Even against a background of conservative education systems, the ever-increasing use of computers in the wider society will inevitably produce changes in the relative valuation of different learning achievements.

From the outset, the topic of "computers for learning" has excited strong passions. Critics (e.g. Chandler, 1992; Sloan, 1984) have been every bit as vociferous as devotees. The language of psychopathology, of phobias, and addictions has been freely used in relation to computer use, albeit with little apparent justification (Crook, 1992, 1994). The use of computers for learning has been seen as depersonalising and alienating by some, as creative and liberating by others.

The discussion of issues of gender, socioeconomic status, and learning disability reflects some of the ways in which computers may exert differential effects upon children's learning; increasing opportunities for some, but marginalising or excluding others. All change serves some interests better than others. Different uses of computers for learning thus not only reflect different theories of learning but also serve different ideological agendas.

Psychological theories of learning have been important sources of inspiration in this field, and the use of computers for learning has provided a test-bed for many such theories. Arguably, none have stood up particularly well to the test thus far, but despite this, computers have secured a place in children's learning just as they have secured a place in almost all aspects of our everyday lives. They have led to, and will continue to lead to, significant changes in both what and how children learn.

Acknowledgement

The preparation of this paper was assisted by a grant from the Leverhulme Trust and by an ESRC Senior Research Fellowship.

References

Blaye, A., Light, P., Joiner, R., & Sheldon, S. (1991). Collaboration as a facilitator of planning and problem solving on a computer based task. *British Journal of Developmental Psychology*, *9*, 471–483.

Bork, A. (1980). Learning through graphics. In R. Taylor (Ed.), *The computer in the school*. New York: Teachers College Press.

Chandler, D. (1992). The purpose of the computer in the classroom. In J. Benyon & H. McKay (Eds.), *Technological literacy and the curriculum*. London: Falmer Press.

Chen, S., & Bernardopitz, V. (1993). Comparison of personal and computer assisted instruction for children with autism. *Mental Retardation*, *31*, 368–376.

Clements, D., & Nastasi, B. (1992). The role of social interaction in the development of higher order thinking in Logo environments. In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving*. Berlin: Springer-Verlag.

Coombs, N. (1989). Using CMC to overcome physical disabilities. In R. Mason & A. Kaye (Eds.), *Mindweave*: *Communication, computers and distance education*. Oxford: Pergamon.

Costa, E. (1991). The present and future of intelligent tutoring systems. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. Berlin: Springer-Verlag.

Crook, C. (1992). Cultural artefacts in social development; the case of computers. In H. McGurk (Ed.), *Childhood social development: Contemporary perspectives*. Hove, UK: Erlbaum.

Crook, C. (1994). *Computers and the collaborative experience of learning*. London: Routledge.

Culley, L. (1993). Gender equity and computing in secondary schools. In J. Benyon & H. McKay (Eds.), *Computers into classrooms*. London: Falmer Press.

De Corte, E., Verschaffel, L., & Schrooten, H. (1992). Cognitive effects of learning to program in Logo. In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving*. Berlin: Springer Verlag.

diSessa, A., Hoyles, C., & Noss, R. (1991). *Computers and exploratory learning*. Berlin: Springer-Verlag.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon.

Elshout, J. (1992). Formal education versus everyday learning. In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.). *Computer-based learning environments and problem solving.* Berlin: Springer-Verlag.

Fawcett, A., Nicolson, R., & Morris, S. (1993). Computer based spelling remediation for dyslexic children. *Journal of Computer Assisted Learning*, *9*, 171–183.

Foot, H., Morgan, M., & Shute R. (1990). *Children helping children*. Chichester, U.K.: Wiley.

Giacquinta, J., Bauer, J., & Levin, J. (1993). *Beyond technology's promise*. Cambridge: Cambridge University Press.

Goldenberg, E. P. (1979). *Special technology for special children.* Baltimore, MA: University Park Press.

Gray, D. (1995). Computer assisted learning and hearing impaired children: Does CAL work? *Journal of the British Association of Teachers of the Deaf*, *19*, 38–46.

Greenfield, P. (1984). *Mind and media: The effects of television, video games and computers*. Cambridge: Harvard University Press.

Hawkridge, D., Jaworski, J., & McMahon, H. (1990). *Computers in Third World schools*. London: Macmillan.

Hawkridge, D., & Vincent, T. (1992). *Learning difficulties and computers*. London: Jessica Kingsley.

Healy, L., Pozzi, S., & Hoyles, C. (1995). Making sense of groups, computers and mathematics. *Cognition and Instruction*, *13*, 505–523.

Holland, J. (1965). Research on programming variables. In R. Glaser (Ed.), *Teaching machines and programmed learning*. Washington, DC: National Education Association

Hopkins, M. (1991). The value of information technology for children with emotional and behavioural difficulties. *Maladjustment and Therapeutic Education*, *9*, 143–151.

Hoyles, C., & Sutherland, R. (1989). *Logo mathematics in the classroom*. London: Methuen.

Hughes, M., & Greenhough, P. (1995). Feedback, adult intervention and peer collaboration in initial Logo learning. *Cognition and Instruction*, *13*, 525–539.

Hutchins, G., Hall, W., & Colbourn, C. (1993). Patterns of students' interactions with Hypermedia systems. *Interacting with Computers*, *5*, 295–313.

Jones, A., & McMahon, H. (1994). The use of the computer as a therapeutic tool for children. In H. Foot, C. Howe, A. Anderson, A. Tolmie, & D. Warden (Eds.), *Group and interactive learning*. Southampton, U.K.: Computer Mechanics Publications.

Jones, A., & Selby, C. (1995). The use of computers for self expression and communication. In D. Jonassen & G. McCalla (Eds.), *Proceedings of the International Conference on Computers and Education*. Charlottesville, VA: AACE.

Kirkup, G. (1992). The social construction of computers. In G. Kirkup & S. Keller (Eds.), *Inventing women: Science, gender and technology*. Oxford: Polity Press.

Light, P., & Light, V. (in press). Reaching for the sky: Computer supported tutorial interaction in a conventional university setting. In K. Littleton & P. Light (Eds.), *Learning with computers: Analysing productive interaction*. London: Routledge.

Light, P., Littleton, K., Bale, S., Joiner, R., & Messer, D. (in press). Gender and social comparison effects in computer based problem solving. *Learning and Instruction*.

Light, P., Littleton, K., Messer, D., & Joiner, R. (1994). Social and communicative processes in computer based problem solving. *European Journal of Psychology of Education*, *9*, 93–109.

Littleton, K., Light, P., Joiner, R., Messer, D., & Barnes, P. (1995). Gender and software effects in computer based problem solving. *Newsletter of the European Association for Research on Learning and Instruction*, December 4–8.

Martin, R. (1991). School children's attitudes towards computers as a function of gender, course subjects and availability of home computers. *Journal of Computer Assisted Learning*, *3*, 187–194.

Mason, R., & Kaye, A. (1989). *Mindweave: Communication, computers and distance education*. Oxford: Pergamon.

Mellar, H., Bliss, J., Boohan, R., Ogborn, J., & Tompsett, C. (1994). *Learning with artificial worlds*. London: Falmer Press.

Mercer, N. (1995). *The guided construction of knowledge*. Clevedon, U.K.: Multilingual Matters.

Mevarech, Z., Silber, O., & Fine, D. (1991). Learning with computers in small groups. *Journal of Educational Computing Research*, *7*, 233–243.

Nathan, M., & Resnick, L. (1994). Less can be more: Unintelligent tutoring based on psychological theories and experimentation. In S. Vosniadou, E. De Corte, & H. Mandl (Eds.), *Technology based learning environments*. Berlin: Springer-Verlag.

Newman, D., Griffin P., & Cole, M. (1989). *The construction zone*. Cambridge: Cambridge University Press.

Newton, P., & Beck, E. (1993). Computing: An ideal occupation for women? In J. Benyon & H. McKay (Eds.), *Computers into classrooms*. London: Falmer Press.

Nichol, J., Briggs, J., & Dean, J. (1988). *PROLOG*: *Children and students*. London: Kogan Page.

Niemiec, R., & Walberg, H. (1987). Comparative effects of computer assisted instruction: A synthesis of reviews. *Journal of Educational Computing Research*, *3*, 19–37.

Olson, D. (1986). Intelligence and literacy. In R. Sternberg (Ed.), *Practical intelligence*. Cambridge: Cambridge University Press.

O'Malley, C. (1995). *Computer supported collaborative learning*. Berlin: Springer-Verlag.

O'Neill, B., & McMahon, H. (1991). Opening new windows with Bubble Dialogue. *Computers and Education*, *17*, 29–35.

Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. New York: Harvester Wheatsheaf.

Papert, S. (1994). *The children's machine*. London: Harvester Wheatsheaf.

Pea, R., & Kurland, M. (1984). The cognitive effects of learning computer programming. *New Ideas in Psychology*, *2*, 137–168.

Pelgrum, W., & Plomp, T. (1993). *The IEA study of computers in education*. Oxford: Pergamon.

Rueda, R. (1992). Characteristics of teacher-student discourse in computer-based dialogue journals. *Learning Disability Quarterly*, *15*, 187–206.

Robinson, B. (1993). Communicating through computers in the classroom. In P. Scrimshaw (Ed.), *Language, classrooms and computers*. London: Routledge.

Saloman, G., Perkins, D., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, *20*, 2–9.

Sardello, R. (1984). The technological threat to education. In D. Sloan (Ed.), *The computer in education*. New York: Teachers College Press.

Scanlon, E., Murphy, P., & Issroff, K. (in press). Collaborations in a primary classroom. In K. Littleton & P. Light (Eds.), *Learning with computers*. London: Routledge.

Schank, R., & Cleary, C. (1995). *Engines for education*. Hillsdale, NJ: Erlbaum.

Schofield, J. (1995). *Computers and classroom culture*. Cambridge: Cambridge University Press.

Scott, T., Cole, M., & Engel, M. (1992). Computers and education: A cultural constructivist perspective. *Review of Research in Education*, *18*, 191–251.

Skinner, B. F. (1965). Reflections on a decade of teaching machines. In R. Glaser (Ed.), *Teaching machines and programmed learning*. Washington, DC: National Education Association.

Sloan, D. (1984). *The computer in education*. New York: Teachers College Press.

Smith, R. (1986). The Alternate Reality Kit: An animated environment for creating interactive simulations. *Proceedings of the IEEE Computer Society*. Los Altos, California, 99–106.

Smith, R. (1991). A prototype futuristic technology for distance education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. Berlin: Springer-Verlag.

Suppes, P. (1966). The use of computers in education. *Scientific American*, *215*, 207–220.

Swettenham, J. (1996). Can children with autism be taught to understand false belief using computers? *Journal of Child Psychology and Psychiatry*, *37*, 157–165.

Torgesen, J., & Barker, T. (1995). Computers as aids in the prevention and remediation of reading disabilities. *Learning Disabilities Quarterly*, *18*, 76–87.

Turkle, S., & Papert, S. (1990). Epistemological pluralism: Styles and voices within the computer culture. *Signs*, *16*.

Yankelovich, N., Meyrowitz, N., & van Dam, A. (1991). Reading and writing the electronic book. In O. Boyd-Barrett & E. Scanlon (Eds.), *Computers and learning*. Wokingham, U.K.: Addison Wesley.

84

HYPERMEDIA AS AN EDUCATIONAL TECHNOLOGY

A review of the quantitative research literature on learner comprehension, control, and style

A. Dillon and R. Gabbard

By virtue of its enabling rapid, nonlinear access to multiple forms of information, hypermedia technology is considered a major advance in the development of educational tools to enhance learning, and a massive literature on the use of hypermedia in education has emerged. The present review examines the published findings from experimental studies of hypermedia emphasizing quantitative, empirical methods of assessing learning outcomes. Specifically, the review categorizes this research into three themes: studies of learner comprehension compared across hypermedia and other media, effects on learning outcome offered by increased learner control in hypermedia environments, and the individual differences that exist in learner responses to hypermedia. It is concluded that, to date, the benefits of hypermedia in education are limited to learning tasks reliant on repeated manipulation and searching of information and are differentially distributed across learners depending on their ability and preferred learning style. Methodological and analytical shortcomings of the literature limit the generalizability of all findings in this domain. Suggestions for addressing these problems in future research and theory development are outlined.

In recent years, the emergence of digital documents has progressed from word-processed text, through stand-alone hypermedia, to the World Wide Web. With each new stage of technological development, the lessons learned from user studies of previous technologies tend to be overlooked. With one eye on the future, many educators and literary scholars are predicting nothing less than a paradigm shift in the manner in which we understand the

Source: *Review of Educational Research*, 1998, 68(3), 322–349.

learning experience and the education process as a result of hypermedia technologies in general and the World Wide Web in particular. For example, Landow (1992) writes: "Electronic linking shifts the boundaries between one text and another as well as between the author and the reader and between the teacher and the student" (p. 33). In a similar vein, Dryden (1994) argues that hypermedia environments can indeed promote the appreciation of literature (and of texts in other disciplines) as they nurture the growth of the learner in intellect and spirit. [Furthermore,] hypermedia has the potential to transform the structure of both classrooms and entire institutions—schools and universities—and to make the teaching and practice of literate thinking and behavior a truly democratic enterprise that respects and serves the needs of both the individual learner and the larger community of learners. (p. 284)

These are but two quotations from a large and expanding literature on hypermedia where this technology is unquestioningly advocated as an advance in educational technology for one or more of the following reasons: (a) Hypermedia enables nonlinear access to vast amounts of information (Nielsen, 1995); (b) users can explore information in depth on demand (Collier, 1987); (c) interaction with the instructional material can be self-paced (Barrett, 1988); (d) hypermedia is attention capturing or engaging to use (Jonassen, 1989); and (e) hypermedia represents a natural form of representation with respect to the workings of the human mind (Delany & Gilbert, 1991).

Such writing is strong on claims but, so far, short on supporting evidence from studies of learners, as some researchers have continually noted. For example, McKnight, Dillon, and Richardson (1991, 1996) argued that the empirical evidence for any educational benefits of hypermedia was not convincing, and recent critiques (e.g., Dillon, 1996) suggest that the breakthroughs promised by hypermedia advocates are more mythical than real. Most telling, perhaps, Landauer (1995) reported that despite numerous published reports on the topic of hypermedia use, he could find only nine studies of human performance with this technology that met even minimally acceptable scientific criteria. Chen and Rada (1996) managed to identify 23 experimental studies involving human interaction with various forms of hypertext up to 1993. These authors adopted less strict criteria for acceptance and counted papers with more than one study repeatedly on the basis of the number of experiments reported (they identified a total of 18 papers); however, even their analysis of effect size showed little real advantage for hypertext over other media in general information tasks.

Focus of the review

In an attempt to advance the arguments about educational hypermedia onto an empirical footing, the present article extends Landauer's analysis into the learning domain and seeks to provide a baseline review of the experimental findings to date on the quantitative effects of hypertext/hypermedia

on learning outcome. Following Landauer's initiative, we sought published studies of hypermedia use and learning outcome that were empirical (based on user data), experimental (here considered as meeting rudimentary scientific requirements for selection, manipulation, and control of variables), and primarily quantitative (although use of qualitative data in parallel did not rule out published studies). Our emphasis was on the *measured* effects of hypermedia usage on learning outcomes, which we defined here as any desirable and demonstrable changes in learner behavior or task performance as a function of instruction or information presentation. Thus, we did not consider directly the expanding literature on hypermedia interface design, where speed and accuracy of location are the primary dependent variables, although studies employing these variables were included where the emphasis on learning and comprehension of material was paramount. For this review, we considered hypermedia to be a generic term covering hypertext, multimedia, and related applications involving the chunking of information into nodes that could be selected dynamically (McKnight et al., 1991).

As such, our focus was deliberately narrow. Studies emphasizing the development process underlying hypermedia applications (included in Chen & Rada, 1996, and Landauer, 1995) and those reporting qualitative learner/instructor responses were not included in the present review. There are many accounts of hypermedia development and design (e.g., Kahn, Peters, & Landow, 1995), but these accounts rarely afford details of subsequent use by learners and thus have little relevance to this discussion. Similarly, it is not our intention to enter the quantitative versus qualitative debate (which we see as uninformative) but to emphasize the quantitative studies of learner performance in order to provide a benchmark. A subsequent review of qualitative studies is both necessary and indeed desirable; at this time, however, we believe that a clear statement of quantitatively measured effects will offer the most insight, and with the literature increasingly being published across disciplines, a synthesis and distillation of findings to date seems necessary. Furthermore, as shown later, the literature on hypermedia has expanded at such a rapid rate (probably more rapid than many of us realize) that reviewing completely the more than 2,000 articles in this domain is beyond any one article.

Research methods

The review concentrated on research findings published between 1990 and 1996 and abstracted or cited in the Educational Resources Information Center (ERIC) database or the PsycLIT database. These databases were selected as representative of the core literature indexes in the areas of education and learning. We found few pre-1990 studies on hypermedia that warranted inclusion, and we wished to avoid criticisms of basing conclusions on 10-year-old technologies that may have significantly evolved. Naturally,

new findings are appearing all the time, and the established databases serve as a filter on the literature in both negative and positive ways; our purpose, however, is to show what is being picked up and categorized in the field (itself an important quality control process), and any slight time lag this involves is indicative of all research domains served by the academic databases.

Multiple searches were performed on each database. Searching on the queries *hypermedia and learning* and *hypertext and learning* yielded 397 citations in ERIC alone from 1990–1995. A second ERIC search using the keyword query *hypermedia.maj. and (learning or (instructional.adj (effectiveness or design))).maj.* resulted in 101 citations that were compared with the first set of results; duplicates were eliminated. A PsycLIT search initiated with the keyword query *(hypertext or hypermedia) and (cognit* or learning or study)* resulted in 63 citations. Each citation was reviewed to determine conformance with the criteria defined earlier. This left a combined list of 97 articles to be reviewed in detail against the selection criteria. After reviewing these, we found 25 that warranted detailed review for this article.

A final round of searches was completed in fall 1996. The ERIC database was searched via the keyword query *hypermedia.maj. and (learning or (instructional.adj (effectiveness or design))).maj*, with the years limited to 1995 and 1996 (to capture new entries); this resulted in 21 citations. The PsycLIT database was searched via the keyword query *(hypertext or hypermedia) and (cognit* or learning or study)*, resulting in six citations. The two lists were combined and duplicates eliminated. The new list provided five additional articles, for a final total of 30 that met all of the study criteria.

Supplementary articles cited by authors reviewed and/or known to the present authors through their own research works were included in instances in which they offered unique perspectives to this main body of work. While this is not a large document set, it represents a substantial increase in the number on which Landauer (1995) based his conclusions (precise overlap is impossible to assess since Landauer did not provide a full list). The final set included only 3 of the 18 articles reviewed by Chen and Rada, the remainder of their set being rejected on the grounds of (a) pre-1990 publication (6 articles), (b) not directly measuring learning outcome (6 articles), or (c) being unpublished in the mainstream literature (Chen and Rada included 3 unpublished dissertations in their set).

It is easier to understand the basis of exclusion by offering an example. We did not include a study by Frey and Simonson (1993) that used a hypercard document on historical costume to investigate the relationship between cognitive styles and media use in a hypermedia treatment. Eighty undergraduate students were given a learning style profile to establish a cognitive style baseline, although the "styles" referred to are more appropriately skills (e.g., analytic skill, memory skill). In addition, they were given a test to establish their level of prior knowledge about the subject matter. After the hypermedia

lesson, students were given a posttest to determine their knowledge of the subject matter. The authors found that knowledge of subject matter did increase significantly after use of the hypermedia treatment relative to pretest scores, although it is not clear precisely what knowledge, if any, students had at the start. While certainly quantitative, few conclusions are capable of being drawn from this study since no controls (in terms of learning environment or student ability) were used. Such studies are typical of published accounts of this technology.

At the outset, it should be noted that synthesizing this literature is an exceedingly difficult task. Chen and Rada (1996) grouped all of their studies under the headings of effectiveness and efficiency, regardless of the variables measured; such an analysis, it is argued here, is impossible to apply meaningfully in this literature. For example, most of the results in the education literature are not significant. Furthermore, as Chen and Rada noted, most of the experimental studies of hypermedia involve unique applications (frequently designed by the experimenters themselves), investigate distinct learner populations (varying in age, skills, experience, ability, learning style, and all combinations therein), assess distinct learning tasks, and quantify learning outcome or process of use differently. As a result, it is frequently necessary to describe in detail the investigative methodology of specific studies. However, the general review can be broken into three major themes, each representing an issue of learning on which groups of researchers have focused directly: (a) comprehension of presented materials, (b) learner control over presentation of material, and (c) individual differences in learning style.

Comprehension is a classic outcome measure of performance and perhaps the strongest test of a learning technology. In these studies, researchers compared hypermedia with other media (e.g., paper) or compared various hypermedia versions of information, and they measured the performance of learners with these tools. The second theme is a process issue relating to the control of presentation, pace, and movement through the information space, a variable that is thought to improve the sense of control learners have over their task and, theoretically, is thought to have positive effects on learning outcomes (e.g., Landow & Delany, 1991). The third issue is a form of individual difference analysis, with the focus on types of learners for whom certain forms of hypermedia might offer specific learning advantages.

Certainly, this division is overly clear cut, several studies measuring components from more than one group. The use of this categorization here is primarily as an aid for the reader. However, the presence of these themes in the literature does reflect an awareness that simple measures of learning are unlikely to provide a complete answer to the complex question of how hypermedia affects learning.

There appears to be no clear adoption of one learning theory in this research. Many of the studies seem to reflect pragmatic rather than theoretical concerns. Exceptions appear in the work of Jonassen and Wang (1993),

who attempted to relate hypermedia use to the formation of cognitive structures in a classic schematheoretic manner, and the work of Jacobson and Spiro (1993), who proposed perhaps the richest theoretical model in this domain—their epistemic beliefs and preferences model—to examine the role of hypermedia in the learning of complex, cross-referenced knowledge. Beyond these articulations of theory, there are standard adoptions of cognitive style and individual difference perspectives on learning, but no other formal learning theories are made explicit in this literature. This issue is examined further in the Discussion section.

Comprehension

When someone reads a text or participates in a class, it is generally assumed that he or she ends this process with some knowledge or information he or she previously lacked. Hypermedia presentation is considered to improve comprehension by virtue of its capability of supporting structured access, rapid manipulation, and individual learner control. Comprehension measures thus seek to estimate this gain in knowledge. However, as several researchers have noted (Dillon, 1992; van Dijk & Kintsch, 1983), there is no universally agreed-upon measure of comprehension, and thus comparisons across studies are rarely straightforward. The studies reported in this section differed in their measures as well as their methods (e.g., some compared paper texts and hypertexts, whereas others compared various hypermedia structures). However, all reported the use of hypermedia environments and relied on experimental methods for their comparisons.

Hypermedia and paper

The majority of experimental findings to date indicate no significant comprehension difference using hypermedia or paper. This appears to be the case for both complex (e.g., essay writing) and comparatively simple (e.g., immediate recall) task measures. However, the experimental designs used and the dependent variables observed make simple descriptions of this conclusion difficult. To aid matters, Table 1 provides an overview in terms of tasks and measures used for all of the experiments discussed explicitly in this section. As noted subsequently, the most frequent finding is one of no significant difference between the media, regardless of the investigative methodologies employed.

It is worth considering the first four studies in Table 1 together. In each, learners were allocated to paper or hypermedia environments and, on the basis of their exposure, tested for comprehension afterward. In all cases, no significant performance differences on comprehension tests were observed. However, despite the general similarities, there were task differences between

Table 1 Comparisons of hypermedia and other media in terms of learner comprehension outcome

Authors	*Comparison basis*	*N*	*Task and duration*	*Dependent variables*	*Result (1/2)*
Aust et al. (1993)	paper	80	Translation over single trial	Proposition recall	*NS*
van den Berg and Watt (1991)	paper and lecture	28 30	Study statistics over 6 weeks	Exam results	*NS*
McKnight et al. (1992)	paper	24	Read for comprehension, single session	Essay quality	*NS*
Becker and Dwyer (1994)	paper	44	2 sessions	Test on material	*NS*
Lehto et al. (1995)	paper	15	Read to learn task (1) followed by reference location task over single session (2)	Time to form topic (1) Time to summarize (1) No. of refs cited (1/2) Time to find references Time to complete task (2)	*NS* *NS* $p<.10/.05$ *NS* $p<.05$
Marchionini and Crane (1994)	paper	20	Essay writing	Number of citations Essay quality	*NS* *NS*
Psotka et al.	paper	10	Visual discrimination tasks, after 30 minutes.	Visual test	$p<.05$
Blanchard (1990)	lecture/lab		Learn MS-DOS over 8 weeks	Mean test scores	*NS*

these studies. Aust, Kelley, and Roby (1993) compared students using one of four learning environments involving permutations of paper or electronic text and their equivalent monolingual or bilingual dictionaries. Eighty undergraduate students who were enrolled in a fifth-semester Spanish language course were randomly assigned to one environment. A 420-word Spanish text (judged to be of moderate difficulty by two experts who identified 65 propositions in the text as units of measurement) was used, and a paper-and-pencil posttest asked students to recall as many of the propositions from the article as they could.

Obviously, the text and the task used in the Aust et al. study were narrow and perhaps unrealistic; however, in a more elaborate test over several semesters (and employing a lengthier digital document), van den Berg and Watt (1991) developed what they termed a *level of abstraction structured text* (LAST) covering introductory statistics and hypothesis testing. LAST documents are hierarchically organized hypermedia in which the frames (pages) form a logical tree, thereby providing a guiding structure for learners.

During the first semester, 28 students were randomly assigned to use the LAST document for six weeks instead of going to lectures, while a control group attended lectures. In the second semester, 30 students were randomly assigned to use the LAST document as a supplement to the lectures, while a control group only attended the lectures. During the third semester, the entire class used the LAST document as their sole instructional source. It should be noted that all students in both the first and second semesters attended lectures for the first 5 weeks. The authors remarked that, on the basis of exam results, there was no consistent difference between the students who used the LAST and those who were in standard lectures. In addition, the level of performance did not differ across the three instructional settings. Hence [there] would appear to be no basis for choosing hypertext over traditional lectures, or vice versa, nor for choosing one instructional use of LAST over any other. (van den Berg & Watt, 1991, p. 123)

Becker and Dwyer (1994) further corroborated these results in their study of undergraduate business students in a beginning course on auditing and computer viruses. The authors developed two treatments, a paper packet and a hypertext program. A pretest ensured a uniform level of background knowledge among students. Two sessions were scheduled. The first covered the computer virus material and was employed to provide students using the hypertext programs with some experience using hypertext. The control group simply read the virus material. The second session covered the auditing material. After completion of both sessions, students completed a posttest. Becker and Dwyer also found no significant difference between the posttest scores of the hypertext group and the paper group.

Using the essay writing method of comprehension assessment, McKnight et al. (1992) also reported no significant difference between graduate students' performance after exposure to material presented in hypermedia or paper.

Using the GUIDE™ application package for hypermedia presentation, these authors had students taking a course in ergonomics study a 10,000-word document on user-centered systems design. Students were randomly allocated to either condition and told to take as many notes as they required with a view to producing a synopsis of the article after the task. After 1 hour, all learners were required to stop reviewing the document and to start writing. A domain expert unaware of the experimental conditions graded the essays.

While the exact tasks and the dependent variables were different in each of these studies, the effect of the task itself was not systematically manipulated in any of them. Since hypermedia is a powerful means of manipulating large amounts of data, presumably tasks that require such actions are likely to be better supported in the electronic domain than on paper. Tackling precisely this issue, Lehto, Zhu, and Carpenter (1995) compared learner performance on a *reference task* and a more traditional *learning task*. After training, 15 graduate students performed two tasks: (a) a *reading-to-learn* task (which required students to browse and comprehend) and (b) a *reading-to-do* task (which required students to find and record which annotations contained information on a topic). Two reading-to-learn questions were to be answered using the hypermedia treatment and two using the paper treatment. Time to form topic, time to answer (summarize), and percentage of relevant references cited by participants were recorded. After completing the reading-to-learn task, the participants were asked to complete 10 reading-to-do questions randomly divided between the hypermedia and paper treatments. Time to find relevant references and percentage of relevant references cited by participants were then recorded.

The authors reported that in the reading-to-learn tasks, paper users provided significantly more correct references than did hypermedia users (although this result was reported as significant only at the $p < .10$ level). Hypermedia users also took slightly more time to form a topic but slightly less total time to answer questions, although neither of these two findings were statistically significant. The authors conservatively noted that, "taken together, these results clearly do not show an advantage of hypermedia over the book for the reading-to-learn task" (Lehto et al., 1995, p. 304), although one might add that the significant effect for correct references rather unambiguously suggests a distinct disadvantage for hypermedia users.

Interestingly, the results of the reading-to-do test showed a statistically significant (this time at the standard $p < .05$ level) advantage for hypermedia on measures of time to complete task and percentage of correct references cited. The authors explained this finding in terms of the more flexible search strategies hypermedia makes possible; it is not clear from their data, however, that strategy differences explain the findings as elegantly as the speed and power advantages of electronic searching, since the *to-do* tasks seem to have been little more than word searches. Regardless of the reason, the task dependency of these results suggests one clear direction for further research:

location of target information in large documents, as opposed to broad comprehension measured after exposure, where hypertext's increased functionality may offer advantages.

It could be argued that the studies reported so far compared the two media for constrained or small information resources, and the pattern of results observed may, in part, reflect this. Where much larger information resources are employed, the advantages of hypermedia might become more obvious, if only because of the advantages of searching and speed of access afforded by this technology. Fortunately, results from a study of an extremely large information resource have been published.

The Perseus Project is a hypercard program consisting of texts (including original full Greek texts of seven authors, partial texts of three more, English translations of all texts, and historical background material), language tools (e.g., Greek-English lexicon), approximately 30,000 images, and reference tools. Included is a path tool allowing users to follow or create and store paths through the material. This hypermedia application uses, for the most part, implicit links, since the designers felt that explicit links represented an editorial act that would inhibit the environment they were creating. Perseus was based at six locations and, over 3 years, 640 students and 20 instructors participated in the field tests. Using the Perseus Project as their hypermedia treatment, Marchionini and Crane (1994) explored what they termed the three important characteristics of learning and teaching in a hypermedia environment: access, freedom, and collaboration.

Obviously, such a resource is a wonderful test case for hypermedia, and there are multiple variables that could be examined. For the purposes of the present review, the major points to note are as follows. Marchionini and Crane (1994) reported quantitative results from two specific studies: The first compared 10 students using the Perseus lexicon and another 10 using a paper lexicon during a translation task, and the second compared 10 students using Perseus to study several Greek plays and a control group that did not use Perseus to cover the same material. The authors reported no significant differences in total time to conduct searches in Perseus or on paper, and, more important, they reported no significant difference between students on traditional measures of critical thinking (essays and translations). However, the number of citations present in student essays was significantly greater for students using Perseus, an observation in line with the Lehto et al. (1995) finding that locating references seems to be enhanced in hypermedia environments.

While some authors have interpreted the lack of difference between the media as a sign of progress (e.g., McKnight et al., 1992, considered their result a triumph for the user-centered design process they had followed, indicating that hypermedia could at least produce results equivalent to paper), two published studies have produced significantly positive comprehension results for hypermedia. Psotka, Kerst, and Westerman (1993) devised

two experiments covering information contained in the Army field manual on recognition of aircraft types. In their first experiment, 10 undergraduate students were given either a paper or a hypermedia copy of the manual and 30 minutes to learn 20 airplane types. Paper manual users were given instruction on how to use the manual to make side-by-side comparisons of aircraft types (pages were folded in half and compared with other pages folded in the same manner). At the end of 30 minutes, a posttest was administered. The hypermedia group outperformed the paper group (mean score: 17 vs. 12.8). This effect was replicated in a further study in which subjects were asked to generalize their knowledge of aircraft types by identifying the same set of aircraft from totally new photographs (a task more akin to comprehension than the original identification/recall task). The researchers attributed their findings to the functionality of rapid access in hypermedia, which enabled learners to develop a better sense of similarities and differences between objects.

However, it must be noted that the two versions of the material were not well matched. The hypermedia tool provided color contrasting, supporting the superimposing of one airplane over another for comparison of shape, apparent motion contrasting (one airplane presented after the other for similar planes, in rapid succession), and similarity clustering of like objects in the visual display. While the results suggest that the digital version can certainly be designed to improve on typical paper versions, the scope for improving the latter was not explored. More than anything else, it can be concluded that, in visual categorization and discrimination learning, the use of animation and superimposition made possible in hypermedia clearly has an important impact on learner performance. However, while hypermedia supports such forms of presentation, they are by no means unique to this technology.

In a more standard comprehension task, Blanchard (1990) examined community college students using a hypermedia system to learn MS-DOS. The objective was to have at least 80% of the students pass the proficiency test after 8 weeks of class. Students were given a pretest at the beginning of the class to determine their level of knowledge. Unfortunately, the control group, which used the traditional lecture/lab method, did not complete the pretest. Blanchard found that, in the four classes using the hypermedia system (vs. the control group), a greater number of students passed the test, and these students had higher average test scores. In the four classes using the hypermedia system, there was a significant increase in the percentage passing and in the average test scores between the pretest and the posttest. Obviously, no pretest/posttest data were available for the controls.

While Blanchard's study was flawed methodologically, the combination of Psotka et al.'s (1993) finding on visual learning, Lehto et al.'s (1995) results with respect to reading-to-do, and Marchionini and Crane's finding on reference citation suggests a strong task dependency to the successful exploitation of this technology. Thus, while it seems that paper offers significant advantages over hypermedia in some comprehension tasks, those tasks (or subtasks)

that involve substantial amounts of large document manipulation, searching through large texts for specific details, and comparison of visual details among objects are potentially better supported by hypermedia.

Structural comparisons of hypertexts

While comparisons of paper and hypermedia suggest that there is no simple answer to the question of which medium is better for learning, the ease with which one may organize and structure material has long been considered a potential advantage of hypermedia presentations. Indeed, some researchers (e.g., Smith, 1994) have argued that hypermedia can model the knowledge structures of experts in a manner that makes their assimilation by learners more likely. The argument appears to be that hypermedia may support the production of more effective pedagogical resources if it is designed in such a way as to model knowledge structures explicitly. By extension, the relatively poor showing of hypermedia in comparisons with paper may be explained by the failure of researchers to design applications in a manner that allows learners to exploit such structures. Three studies have examined the precise effects of various hypermedia forms on learning, and their results are summarized in Table 2.

Jonassen and Wang (1993) argued that hypermedia's support of structural mapping would lend itself ideally to helping novices acquire an expert's representation of a subject domain (a view termed the "naive associationist" model of hypermedia by Dillon, 1996). These authors tested this theory on preservice teachers in an education program where their task was to learn about the classroom use of hypermedia as a new instructional technology. The authors developed three test instruments, consisting of 10 questions each, covering the following aspects associated with structural knowledge: (a) relationship proximity judgments, (b) semantic relationships, and (c) analogies. An additional multiplechoice test was developed to assess student recall of the information contained in the hypermedia. These test instruments were then used in three studies examining the effects of various hypermedia structures on learning.

In the first experiment, a graphical browser, a pop-up window (both of which provided explicit information about the structure of the document and the nature of the links), and a control hypermedia (where no extra information was provided) were compared. Of the four dependent variables measured (recall, relationship, proximity, and analogy subscales), only the recall variable approached statistical significance ($p < .07$), and this was in favor of the control group.

In the second experiment, Jonassen and Wang tested the value of generating a classification of the link by the students. One hundred twelve students were divided into three groups (generative, pop-up, and control). In this experiment, the link relationships were not displayed to the generative or

Table 2 Comparisons of various hypermedia forms in terms of learner performance outcome

Authors	*Comparison*	*N*	*Task and duration*	*Dependent variables*	*Result*		
Jonassen and Wang (1993)	1) Browser v pop-up menu v. control	98	Study for comprehension over a single trial	proximity judgments	1 *NS*	2 *NS*	3 *NS*
	2) User classified link v. controls	112		semantic relationships	*NS*	*NS*	0.01*
				analogies	*NS*	*NS*	*NS*
	3) Semantic net v control	48		recall	*NS*	*NS*	*NS*
Jacobson and Spiro (1993)	HyperTnedia v CBT	39	Comprehension over 4 days	Recall (1/2) Essay writing (1/2)	$p < .01$* *NS*		$p < .05$* $p < .01$
Tripp and Roby (1990)	Presence/absence of advanced organisers and visual metaphors	60	Language leaming in 15 minutes	Recall test scores	$p < .05$		

pop-up groups; instead, students in these groups were required to classify the link themselves. The rationale here was that forcing learners to attend to the link type would enhance their perception of the semantic structure of the document. If, after two tries, the students were unable to classify the link, the program provided the link type and proceeded to a new node. The authors observed that, 'as in Experiment 1, control group subjects were better able to recall information, as they were less distracted by the structural knowledge activities. Neither structural strategy (generative or pop-up) produced any increase in structural knowledge' (p. 5).

Jonassen and Wang's third experiment involved 48 graduate students divided into two groups. The authors argued that actively engaging learners in their own learning activity should enhance performance. As a result, the treatment group was required to construct a semantic network of ideas from hypermedia material. They were provided a tool with which they already had experience to accomplish this task. A control group was given the task of simply studying the material in the hypermedia. The treatment group performed significantly better than the control group only on the relationships judgment task, suggesting (perhaps) that learners need to become focused on structural relationships to acquire structural knowledge. The authors noted that "merely browsing through a knowledge base does not engender deep enough processing to result in meaningful learning" (p. 6), and they concluded that their initial assumption of the assimilation of expert knowledge being easier in a hypermedia environment was not supported. Perhaps most insightful is the authors' contention that hypermedia might work better as an information retrieval interface than as a learning enhancing tool, in line with the findings of Lehto et al. (1995).

While Jonassen and Wang's results cast some doubts on the simple view of knowledge transfer being enhanced through structural manipulations in hypermedia presentations, it could be argued that the type of knowledge that one seeks to learn is a crucial variable and that the three Jonassen and Wang studies might not have used the most appropriate knowledge type. Jacobson and Spiro (1993) explored such an issue by examining students using hypermedia to learn both complex and ill-structured knowledge. Three environments were developed: two control conditions (minimal hypermedia and computer-based drill) and on experimental treatment (full hypermedia). It was hypothesized that the control groups would achieve higher scores on the memory tests of factual knowledge, while the full hypermedia users would have higher scores on the transfer test. The authors also hypothesized nonspecific user preference differences.

Thirty-nine freshman and sophomore students were given a pretest to determine prior knowledge and then randomly assigned to one of the three conditions. The material covered was characterized as being conceptually ill structured as a result of its multidisciplinary, diverse, and dynamic content. Learners engaged the technology for four sessions, one per day, with testing

of their performance on Days 2 and 4. Testing involved recall of factual material and a written essay. The control groups were significantly more effective and efficient in acquiring factual knowledge, as evidenced by their scores on both recall tests. However, the experimental group outperformed the control groups in the problem-solving essay on Day 4, although no significant difference was observed between groups on their Day 2 essays. The authors suggested that this pattern indicates the task dependency of hypermedia use; it seemingly is of little use (and may even be disadvantageous) for abstraction of factual knowledge (although it is not clear why this should be so theoretically), having more useful application for learning tasks involving synthesis of complex material.

Both of the aforementioned investigations alluded to the dynamics of comprehension in hypermedia environments. The tasks were reasonably complex, but, despite numerous samplings of learning measures, few significant differences emerged. What sets these two studies apart is their explicit articulation of theoretical issues, albeit two very different theories. Jonassen and Wang clearly adopt a structuralist perspective that is seductive but limits the analysis of comprehension to one based on form more than content. Jacobson and Spiro's position is more subtle and allows for a more complex model of learning that hypermedia may support. If progress is to be made in this area, it is likely that the latter authors' approach is the most compelling (an issue we return to later).

Using a comparatively simplified learning task, Tripp and Roby (1990) concentrated on the effects of hypermedia on acquisition of Japanese words. They developed four hypermedia test environments that were intended to manipulate the cognitive processing of learners by crossing the use or absence of an advance organizer for the material with the presence or absence of a visual metaphor about the organization of the database. However, the details of these treatments were somewhat vague, and the visual metaphor seems to have been a background graphic on the screen. The authors hypothesized that students would learn the fewest Japanese words in an unstructured hypermedia environment that lacked organizers or metaphors, since these devices should free up limited cognitive resources for learning.

Sixty undergraduate and graduate students were randomly assigned to one of these environments with the task to learn as many Japanese words as possible in 15 minutes (a task that appears to be lacking in ecological validity). A paper test consisting of multiple-choice and recall questions was given at the end of the session. Learners who were exposed to both the metaphor and the advance organizer actually did significantly worse than either group that had only one of these treatments, and they performed only slightly better than the group with neither of these treatments, confounding the authors' views of the likely cognitive benefits of such treatments. Tripp and Roby concluded that "two types of orienting devices activated conflicting mental models of the lexicon" (p. 122); therefore, while orienting aids such as advance

organizers or metaphors may be of some benefit in the learning task, more is definitely not better.

Hypermedia and other electronic media

As well as considerations of the best form of hypermedia, there has been interest in comparing hypermedia with other electronic media. Saga (1992) explored the effects of hypermedia and video environments on learner's interest and comprehension. A hypercard-videodisc program from a video presentation titled *Great Authors Who Lived in Bunkyo* was adapted, and three subject groups were recruited: 57 college students, 20 audiovisual librarians, and a control group of 17 college students. All subjects were given a pretest measuring their prior knowledge about the subject. The two treatment groups participated in an hour-long presentation that included a lecture, the hypercard-videodisc presentation, and the video. The control group only watched the video. All three groups showed a significant increase in knowledge in the posttest results. However, the control group outperformed the other two groups in a test of factual knowledge. In a conclusion fitting for this entire section of the review, Sage noted that this finding is "impressive enough to warn against excessive expectations from hypermedia" (p. 186).

Summary of comprehension findings

The results on learner comprehension from hypermedia are, at best, inconclusive, but the weight of evidence points to hypermedia being suitable mainly for a limited range of tasks involving substantial searching or manipulation and comparison of visual detail where overlaying of images is important. In short, the evidence does not support the use of most hypermedia applications where the goal is to increase learner comprehension (however measured). Evidence from studies of hypermedia structural variables suggests a particularly limited knowledge base in terms of how best to organize information in a digital form that exploits the cognitive capabilities of learners to link and organize new information.

Learner control

If hypermedia use does not lead directly to gains in comprehension, it might be argued that the medium's advantages really lie elsewhere (i.e., hypermedia's effects are mediated by other variables). For example, many authors have claimed that the capability of digital technology to enhance learner control over the pace and detail of information delivery has a positive effect on learning (Dryden, 1994; Landow, 1992; Landow & Delany, 1991; but see Hooper & Hannafin, 1988, and Relan, 1991, for alternative viewpoints). Since one of the advantages of hypermedia is precisely the control it allows users to have over their access to information, changes in learning outcome might

be best observed under conditions in which learner control is affected. Five published studies have examined this issue, and we examine these studies in turn.

Control of access and learner performance

Manipulating the learner control variable is not a straightforward matter. Most researchers attempt to do this by creating various presentation formats that vary the means by which the learner may manipulate the information displayed. In this way, a hypermedia application with multiple links and/or a graphical browser that offers selectable links is seen as offering more control to the user for access to material than a simple text file that affords only screen scrolling or page turning. While objections may be raised in terms of the validity of such operationalizations of learner control, and it is certainly the case that no fixed scale of controllability is being used here, it is likely that such treatments offer at least one plausible, if partial, manipulation of learner control.

A typical example of such an investigation was reported by McGrath (1992), who investigated the role of learner control in reducing misconceptions during learning. One hundred three undergraduates in a teacher preparation course were randomly assigned to one of four test environments: hypermedia, paper, menudriven electronic pages, or no menu electronic pages (where subjects had no choice but to go on to the next page). A pretest to determine prior knowledge was administered, classifying each student as either high or low in ability and highlighting any previously held misconceptions held by the student on the subject matter. A posttest was administered to determine student achievement.

McGrath hypothesized that the increased control of the hypermedia and menu-driven pages would positively affect learning. However, no significant differences were observed between conditions. Comparisons between learners of high and low skill across all variables were not statistically significant either. Learners made nonsequential responses in both computer environments, which may be interpreted as a sign of manifest control; however, both the high- and low-skill learners made more such responses in the menu environment than in the hypermedia environment. While these findings cast doubts on the value of control on certain learners, there is another possible interpretation (i.e., the increase in user control that is assumed to occur in hypermedia environments might be false).

In a further examination of learner control in electronic environments, Welsh, Murphy, Duffy, and Goodrum (1993) devised three experimental conditions: (a) no link type (NLT), in which all link icons were the same and did not provide any cues as to the information behind the link; (b) link type (LT), in which different link types indicated information types; and (c) submenu (SM), in which a link led to a submenu with elaborations listed. The authors predicted that NLT users would manifest random exploration

of the database, whereas LT and SM users would target their exploration, following or rejecting links in their search. They also predicted that NLT users would experience frustration and follow fewer links than LT or SM users and that high link density on a given page would increase visual noise and disrupt the learner's ability to read from the screen. A 2,000-word core hypermedia database with 320 elaborations was created, and 108 undergraduates were randomly assigned to one of the following six environments: NLT-high density, NLT-low density, LT-high density, LT-low density, SM-high density, or SM-low density.

Learners were tested in groups of 20 for about 30 minutes. A comparison of link type and time on task failed to demonstrate any statistical significance, as did the issue of link density. The authors reported that the NLT condition did stimulate random exploration (although it is not clear how they measured the random variable). Contrary to predictions, NLT users did not experience frustration or become discouraged from the lack of information. The authors concluded, rather generally, that "designers of hypermedia learning environments will have to make tradeoffs in creating systems that promote exploration and in creating systems that have high usability" (p. 33).

Lanza and Roselli (1991) manipulated learner control in their study of student achievement in either computer-aided instruction (considered low learner control) or hypermedia (high learner control) environments. Sixty undergraduate students, with no experience in either computer-aided or hypermedia instruction, were randomly assigned to one of the language learning environments. Unlike most studies of hypermedia use, the learners in this study were observed over an extended period. Covering the same material, the students spent 4 hours per day for two weeks in each learning environment (a trial period was allotted so that the students would become familiar with the learning environment). A 10-question test was given at the end of the 2-week period to measure achievement. The authors reported greater variance in the scores of the hypermedia users, leading them to suggest that hypermedia may not be effective for every learner.

Quade (1993) also examined the role of learner control in a comparison of a computer-assisted instruction tutorial allowing only linear movement and a hypermedia tutorial with a graphical map. Hypermedia users had the option of moving linearly through the material or using the graphic map to select a topic out of sequence or return to the main menu. Both tutorials provided an immediate feedback loop for learners.

Seventy-six undergraduates were randomly assigned to one of the two treatments, and none had any prior classroom work in the subject area. Each student's ability was ranked by her or his professor based on observation and test performance over the previous 8-week period. A 30-question pretest was administered, and students not familiar with the Macintosh platform were given a 2-hour training session. At the end of the experiment, the pretest was readministered as a posttest. According to Quade, learner

control over the number of screens viewed was not found to be a factor in overall subject performance. Furthermore, no significant difference in overall performance (as measured by the posttest) was observed.

If learner control on its own is insufficient or difficult to manipulate, its combination with other variables might offer some clues as to the relative merits of this technology. In a study of learner control and the advice available to learners in a hypermedia environment, Shin, Schallert, and Savenye (1994) manipulated amount of learner control (free access to all contents or limited access through links), presence of advisement, and level of prior knowledge. In this experiment, advice involved recommendations on which sequence to follow in the hypermedia and visual aids for locating oneself in the hypermedia environment. Manipulating advisement and access combinations in a 2×2 design, the authors assigned 110 second graders to work on a lesson concerning food groups. A pretest assessing prior knowledge was administered to each student. Students were randomly assigned to one of the four combinations. After completion of the lesson, a posttest was administered. The same posttest was repeated 1 week later.

Interestingly, the authors reported that students in the limited-access treatments answered more questions correctly than did the free-access students (the authors noted that these results differed from the Lanza and Roselli findings reviewed earlier). Access had no significant effect on students with a high level of prior knowledge, but students with a low level of prior knowledge were more successful in the limited- than free-access environments, suggesting an interesting interaction.

Advisement had no significant effect on learning. It did, however, affect time-in-lesson; students with low levels of prior knowledge finished the lesson quicker when they were in one of the no advisement environments (this result seems to have occurred because they quit programs without knowing that there was more information to cover). The free-access/no advisement environment was difficult for all students but especially those with low prior knowledge. Shin et al. recommended that limited access be used to improve young students' achievement, particularly if they have limited prior knowledge. In a conclusion that seems typical of all educational theories, Shin et al. stated that "learners with different levels of prior knowledge require different kinds of instructional approaches" (p. 45)

Conclusions on learner control and hypermedia

With its embodiment of structure and linked information nodes, hypermedia is considered to offer users far more control over access and exploration. Obviously, control can be manipulated in multiple ways, and the degree of control any one application embodies is difficult to measure; most researchers manipulate this variable through the provision of selectable links and paths. Different students seem to react to this increased control differently, with

lower ability students manifesting the greatest difficulty in exploiting it to their advantage. As a general characteristic of hypermedia environments, the ability to control pace and delivery of information, even when coupled with selection advice, appears insufficient to affect learning outcomes significantly for all but high-ability learners.

Individual differences among learners

While the search for learning gains based on media characteristics or interface features has been largely unsuccessful, a third line of research has focused more directly on characteristics of the individual learner. There have been hints of the importance of learner variables such as level of prior knowledge, for example, in the work of Shin et al. (1994), and experience with hypermedia has been controlled in some studies (e.g., Lanza & Roselli, 1991). Extending this consideration of learner variables logically, several authors have proposed a variety of individual differences among learners as crucial mediating variables in explaining the effect of hypermedia on learning outcome.

Typically, ability is considered a major determinant, but in this literature learning style (conceptualized as reflecting one's distinct *approach* to learning) is frequently postulated as an important variable. Learning styles reflect a learner's position on a continuum running between extreme traits such as holistic and analytic, verbal and spatial, reflective and impulsive, or exploratory and passive. Each trait is seen as having advantages in certain situations, but individuals are supposedly characterized as being predominantly at one end of the continuum or the other.

It is not difficult to envisage how such variables might be important and why they are tempting to use in this domain. The flexibility of hypermedia technology renders it a strong candidate for tailoring the presentation of information to suit particular learners; thus, any empirically demonstrated relationship between such an individual difference variable and effects of hypermedia instruction might offer an explanation for the mass of nonsignificant results to date.

In the present section, 10 published studies on the learning style differences in using hypermedia are reviewed (see Table 3). As shown in Table 3, individual differences have been studied in both large and small samples of learners, and significant interaction effects have frequently been observed. However, not all individual difference variables have proved insightful, and in this section a broad distinction is made between general ability and specific cognitive style differences as they pertain to hypermedia use.

Learner level and hypermedia use

Obviously, the greatest source of individual difference between learners is their general intelligence or level of ability. According to Dillon and Watson

Table 3 Analysis of individual differences in learners and performance with hypermedia

Individual difference	*Author*	*Task*	*N*	*Interaction*
Ability	Recker and Pirolli (1995)	Learn LISP	16	$p<.05$
	Higgins and Boone (1990)	History learning	40	Claimed
	Repman, Willer, and Lan (1993)	Collaborative learning with hypermedia	118	$p<.05$
Field dependence	Lin and Davidson (1994)	History learning	139	*NS*
	Liu and Reed's (1994)	Language learning	63	*NS*
	Jonassen and Wang (1993)	Comprehension of educational material	112	*NS*
	Stanton and Barber (1992)	Process control mastery	20	*NS*
Passive'active	Lee and Lehman (1993)	Comprehension of scientific material	167	$p<.05$
Deep/Shallow	Shute (1993)	Electronic circuits analysis	309	$p<.05$
	Beishuizen, Stoutjesdijk, and van Putten (1994)	2 studies of comprehension of psychology textbook	1) 48 2) 42	$p<.01$ $p<.01$

(1996), meta-analyses of individual differences studies indicate that general ability is the single best predictor of performance on most tasks, and such findings have relevance to all forms of human-computer interaction as well. In the educational literature, general ability has been used as an independent variable in several studies of hypermedia use.

Focusing on the strategies learners use when studying instructional materials prior to problem solving, Recker and Pirolli (1995) developed two instructional environments (hypermedia and a nonhypermedia electronic text) covering the same material (Lisp programming). These authors were interested in the effect on learning of adding elaborations to the instructional material. However, noting that previous studies of the elaboration variable had failed to show advantages, Recker and Pirolli further manipulated that variable by providing the learner with the option of viewing or not viewing elaborations.

Sixteen students with minimal programming skills were recruited for this experiment. During the introductory phase, students worked through four lessons followed by a problem-solving session on a computerized Lisp tutor. The performance of each student on the last set of problem-solving exercises was used to determine an ability or *level of knowledge* measure. In the target phase of the test, students were randomly assigned to either the hypermedia or control environment. They progressed through four more lessons before another problem-solving session on the Lisp tutor. The number of errors occurring during the problem-solving session was used as a measure of student performance.

The authors reported no significant difference between environments. Using their ability measure to distinguish students indicated that the hypermedia environment was mostly beneficial to high-ability learners and that low-ability learners' performance decreased after using the tutor. Recker and Pirolli argued that the lower performing subjects were not able to take advantage of the hypermedia and, in fact, may have been overwhelmed by the amount of learner control it provided (echoing an earlier finding of Lee & Lehman, 1993).

Specifically examining lower ability students, Higgins and Boone (1990) explored the effects of hypermedia study guides on students diagnosed with learning disabilities, remedial students, and "regular" students. The study guides were designed to cover material on the history of Washington State in 10 chapters; there were links throughout each chapter to more detailed information, and a set of study questions followed at the end of each section. The study guides were designed to provide increased instruction time for students without increasing the teacher's instructional load.

Forty ninth-grade students (10 students with learning disabilities, 15 remedial students, and 15 regular students) were randomly assigned to three conditions: lecture, lecture with study guide, and computer study guide only. A pretest was given to all students to establish their base level of knowledge about the subject. The material covered was the same in all three groups. Students in the lecture-only group had worksheets with the same information and questions contained in the hypermedia study guides. Daily quizzes were administered, and a posttest was given at the end of the course. In addition, two weeks after the end of the course, the posttest was given again to measure retention.

Unfortunately, the authors reported many differences in daily quiz scores without alluding to their statistical significance or indicating how many such differences might be found by chance alone; in general, however, regular students significantly outperformed students with learning disabilities in posttest and retention tasks as expected. Even though no significant treatment effects were observed, the authors stated, for reasons that are unclear: "There is evidence that the computer study guide treatment was as effective as the lecture condition" (Higgins & Boone, 1990, p. 535). Oddly, the authors

reported that, in their pretest, significant differences between all student groups were observed in the expected direction (regular students outperformed remedial students, who in turn outperformed students with learning disabilities), except for the computer study guide group. This suggests that the least learning disabled of the learning disabilities group were allocated by chance to the computer condition, suggesting even further caution in interpreting these data.

Higgins and Boone then performed a second study using the same computer study guide to determine whether it could bring five failing students closer to a passing grade. Taking these students' performance in Study 1 as a baseline, they exposed these students to 10 additional computer study lessons and measured their performance again. Results showed that the students raised their grades on the daily quizzes, the posttest, and the retention test (mean increases: 41% to 52%). Without resorting to inferential tests, the authors concluded that the hypermedia treatment was useful in raising the grades of these lower achieving students; without control measures, however, these findings indicate only that improvements are gained by having failing students read and review instructional material again.

Horton, Boone, and Lovitt (1990), using the materials and test instruments just described, studied four students with learning disabilities. In this experiment, control questions were included (questions that were found on all three tests [pretest, posttest, and retention test] but were not covered in the study guide). The authors found that there was a significant overall increase in correct answers between the pretest and the posttest and between the pretest and the retention test; there were no significant differences in scores on the computer items between the posttest and the retention test, suggesting a clear immediate learning effect and good retention. As expected, no effect for control items across the pretest, posttest, and retention test was found. The authors noted that this group of learning disabled students "demonstrated significant improvement on computer questions, but not on control items, . . . and those effects were maintained over a four-week retention period" (p. 129). It is not clear how best to interpret such findings, since the necessary control of nonhypermedia study guides was not used for comparison. At best, the results show, once again, that studying material can improve performance.

Repman, Willer, and Lan (1993) investigated whether novice and low-ability students would benefit from hypermedia-based instruction in a collaborative learning environment. In particular, they wanted to determine whether working alone was significantly different from working in pairs and, in the case of paired working, whether the composition of the pair (in terms of ability levels) was important.

These authors developed a hypermedia-based instructional unit. One hundred eighteen students enrolled in a computer literacy course were randomly assigned to work individually or in pairs during a 50-minute class. A

paper-and-pencil test covering the material was administered the next day. The authors reported no significant effect for working alone or in a pair. However, magnet students, working alone or with a peer, outperformed all others. Nonmagnet students did significantly better when paired with a magnet student than when working alone. Perhaps most worrying, the magnet students working with nonmagnet students scored approximately one standard deviation lower than when working alone or with another magnet student. Repman, Willer, and Lan concluded that students do approach hypermediabased instruction differently and that these differences "vary greatly across social context as well as within each group" (p. 294), suggesting specific further avenues for research.

While it seems that the ability differences among learners certainly account for some of the variance in performance with hypermedia, other individual differences have been proposed as explanatory. Learning style dimensions are considered independent of ability (although correlations between some of these dimensions and ability have been reported; see Dillon & Watson, 1996), and they remain popular since they may offer the clearest indication to educators of how hypermedia interventions can best be targeted at specific learner populations.

Learning style: field dependencelfield independence

A popular source of individual differences seems to be the cognitive style construct of field dependence (FD) and field independence (FI), generally considered to represent differences in preference to attend to specific issues or to rely on context. Lin and Davidson (1994) hypothesized that structured hypermedia would provide an organizational aid to learning that is likely to be differentially useful to FD and FI learners. These authors identified five types of linking structures commonly used in hypermedia (linear, hierarchical, hierarchical-associative, associative, and random) and investigated whether performance was significantly predicted by linking structure, field dependency/independency, and their interaction. They then developed five experimental treatments, each using one of the five linking types, covering material on the Tianamen Square incident. One hundred thirty-nine undergraduates were given the Group Embedded Figures Test to determine cognitive style and were then randomly assigned to one of the five treatment groups. Consistent with other findings, FI learners outperformed FD learners, regardless of environment. The authors observed no difference in performance between learners with the same cognitive styles as a function of linking structure and reported that the "performance of these subjects cannot be predicted by the interaction of linking structure types and cognitive style" (p. 459).

This style dimension was also investigated by Liu and Reed (1994), who asked two questions: (a) What is the relationship of learning style and learning

strategy? and (b) What types of media, tools, and learning aids are preferred by learning style groups? The authors developed a hypermedia (incorporating text, sound, video, and hypermedia links) treatment that emphasized meaningful use of vocabulary in the proper context. The treatment consisted of four subprograms, each containing 20 highlighted vocabulary words. When the highlighted word was selected, the learner was provided with the following options: (a) definition, (b) part of speech, (c) sentence examples, (d) video context, and (e) relationship of the word to other words. Also included was the capability to take notes, on-line help, index tools, and navigation tools.

Sixty-three international students in an intensive English program were given the Group Embedded Figures Test to assess their cognitive style (FD, FI, or field mixed). At the completion of the session, each student took an achievement test. The authors reported that, in the area of learning patterns, FD learners used the various features of the courseware significantly more than FI or field-mixed learners, but no significant results on outcome were observed for the relationship between learning style group and media access or use of learning aids, dictionary, or background information.

Jonassen and Wang (1993) also investigated learner styles as part of their study described earlier. They reported a statistically significant relationship between FI learners and the recall and relationship variables measured in their posttest, supporting recent suggestions that the FI construct more truly measures intelligence than any meaningful cognitive style (Eysenck, 1990). According to the authors, FI learners "were the only learners able to successfully use structural cues to acquire more structural knowledge information. . . . It is likely that field independent learners are better hypermedia processors, especially as the form of the hypermedia becomes more inferential and less overtly structured" (p. 7).

In a study that encompassed all three of the areas covered in this article comprehension, learner control, and learning style), Stanton and Baber (1992) explored the hypothesis that giving learners greater control over their learning should increase the effectiveness of computer-based training. Twenty subjects interacted with a training system on process control that consisted of eight modules. Learners could repeat modules or choose not to interact with a module if they did not feel they required that piece of training. After completion of the training, subjects moved on to the *process control* plant task. The decisions made by the subjects, as well as the status of the plant during the simulation, were recorded, enabling plant "output" to be recorded as an index of task performance. One week later, subjects again interacted with the task (without further training), yielding a measure of the degree of training retention. In addition, the amount of time spent in training, the time spent on each instruction module, and the time spent in practice modules were recorded.

Prior to testing, the subjects were administered an embedded figures test to determine cognitive style. A postexperiment questionnaire was used to

determine their approach or strategy to the training environment. After the experiment, the authors divided the subjects into field dependent/independent groups and learning strategy groups (top down, bottom up, sequential, and elaborative) based on their node visiting sequences. The authors reported significant differences between the cognitive style groups in terms of number of modules visited and completed and time spent in the training sessions; there was no significant difference in transfer task performance. The authors noted that the relationship between field dependence/independence and learning strategy was not conclusive and admitted that the strategies they described might simply be artifacts of their method rather than representations of true user differences, further demonstrating the need for greater methodological rigor in this domain and confirming the general lack of insightful outcomes of research based on the field independence/dependence style construct.

Learning style: passivelactive learners

A learning style dimension of possible direct relevance to hypermedia use is the passivity/activity of the learner. Since hypermedia may support greater direction by the instructor of material to follow, there are grounds for believing that learner passivity/activity may interact significantly with successful learning from this technology.

Lee and Lehman (1993) investigated this construct in terms of control of information delivery in hypertexts covering the topic of DNA and protein synthesis. While both presentations had explicit and implicit buttons linking information, only one involved cuing. In the cuing presentation, students were alerted that more information was available if they chose to view it. One hundred sixty-seven undergraduates were evaluated to determine a prior level of knowledge (a score of above 80% disqualified one learner from this test). The remaining students were given a Passive Active Learning Scale evaluation to classify their learning style. Learners are classified as active if they exhibit curiosity, initiative, and a wide focus while selecting information on their own, and they are classified as passive if they select only information overtly provided and demonstrate indifference, dependence, and a narrow focus. Those who fall between these two categories are termed neutral learners.

The authors reported that active learners outperformed the others on all three dependent measures and that passive learners were outperformed by neutral learners, suggesting some validity to the active/passive construct as a predictor of performance. Cues had no significant effect on active learners, but passive learners showed a significant increase in the achievement variable in the explicit cue condition, suggesting that passive learners may react to prompts for them to follow links. Lee and Lehman advised that designers take such individual differences into account.

Shute (1993) developed two environments: an application environment that provided immediate feedback (in the form of the rule) for correct or incorrect responses and an inductive environment that required learners to discover the rule that determined the correctness or incorrectness of their answer. Three hundred nine subjects were randomly assigned to one of the environments, and two pretests assessed prior subject knowledge. At the end of the trial, a series of four posttests was administered to measure acquired knowledge. Shute hypothesized that learners who manifested exploratory behaviors (i.e., were active) in their performance data would demonstrate higher scores on the posttest if they learned in the environment that allowed exploratory behavior (inductive), while less exploratory (passive) learners would benefit from a more structured environment.

Shute reported that, in the area of learning outcome (percentage correct on the four posttests), neither environment showed a significant advantage over the other. However, high procedural exploratory behaviors were significantly associated with poor outcomes, and the proportion of time allocated to reading definitions was a significant positive predictor of learning outcome. Finally, a significant interaction involving procedural (but not declarative) exploratory behavior and learning environment predicted learning outcome. Shute noted that when learner style is correctly matched to learning environment, learner performance may increase, but a systematic manipulation of this variable a priori would lend stronger support to the exploratory behavior dynamic.

Learning style: deep/shallow processors

Deep and shallow processing refers to the degree of structural or surface analysis and metacognition learners typically manifest in response to new information. Learners can be divided into deep processors who have learning strategies that relate and structure information actively and surface processors whose learning strategies are more passive, usually centered around memorization and rehearsal of information.

Beishuizen, Stoutjesdijk, and van Putten (1994) developed two treatments to test the hypotheses that students with these learning styles profit differently from different types of instructional materials and that increasing task constraints will affect students with different learning styles unequally. In the first experiment, 48 students were given an Inventory of Learning Styles (ILS) evaluation to determine whether they were deep or surface processors, after which they completed a task involving preparation for an examination. Students were randomly assigned to one of three conditions where they received hints about study strategies, content structuring help, or neither. At the end of the study period, students were administered a posttest to determine their knowledge acquisition. No significant main effects were observed

for processing style or hypermedia treatment in either process or outcome measures; however, the authors identified interactions suggesting that content structuring help was detrimental to deep processors but helped surface processors. The authors noted that hypermedia materials are suited for those students who "tend to primarily rely on structures they create themselves" (p. 164).

In the second experiment, students were given a task that emphasized selection of information, and the authors further distinguished learners in terms of *self-regulation* on the basis of a subscale from the ILS. The authors hypothesized that surface processors would make more use of the available text-related guiding tools in their hypermedia assignment, while deep processors were expected to use the map function to select text units in the hypermedia. Forty-two students who had been given an ILS evaluation took part in this experiment using the materials from the first experiment. The authors reported that students who combined self-regulation with deep processing and students who combined external regulation with surface processing performed better than did students with complementary combinations of regulation style and processing style. The results of both experiments were interpreted as confirming that surface processors are less comfortable in the hypermedia reading environment.

Conclusions on learner style

The interaction of learner style in the use of hypermedia offers perhaps the beginning of an explanation for the generally conflicting results in the literature comparing hypermedia and nonhypermedia learning environments. However, the concept of learner style has many meanings, and it is not always clear how, for example, deep processing self-regulators are similar to high procedural exploratory learners. The cognitive style distinction of field independence/dependence remains popular, but, as in most applications to new technology designs, it has failed to demonstrate much in the way of predictive or explanatory power and perhaps should be replaced with style dimensions that show greater potential for predicting behavior and performance.

What seems to be conclusive is the fact that high-ability learners will perform better than low-ability learners, regardless of the medium of instruction, but that hypermedia applications can offer techniques (e.g., explicit cuing) that can help the less able student perform better. Obviously, this area needs much more research to yield the form of evidence that can drive design or exploitation of the technology, but it does suggest that the use of hypermedia in education should be based on appropriately designed technology aimed at specific learners if any significant benefits are to be obtained.

General conclusion

So what are we to conclude from the studies reviewed in this article? Clearly, the benefits gained from the use of hypermedia technology in learning scenarios appear to be very limited and not in keeping with the generally euphoric reaction to this technology in the professional arena. The experimental evidence to date suggests three broad conclusions.

(1) Hypermedia affords the most advantage for users in specific tasks that require rapid searching through lengthy or multiple information resources and where data manipulation and comparison are necessary. Outside of this context, existing media are better than or as effective as the new technology.
(2) Increased learner control over access is differentially useful to learners according to their abilities. Lower ability students have the greatest difficulty with hypermedia.
(3) The interaction of learner style in the use of various hypermedia features offers perhaps the basis of an explanation for the generally confusing results in the literature comparing hypermedia and nonhypermedia learning environments. Specifically, passive learners may be more influenced by cuing of relevant information, and the combination of learner ability and willingness to explore may determine how well learners can exploit this technology.

From these broad conclusions, it can be inferred that the value of hypermedia in pedagogy is limited. Since hypermedia is ultimately a form of information presentation, there should be no real surprise here. That manipulating form of delivery produces mixed results is a reflection of the gaps in our knowledge of how best to design media, and since most educators are fully aware of the multiple forces that shape learning outcome, we should not pin undue hope on any technology of presentation yielding major breakthroughs in terms of education outcomes.

However, in tasks that involve multiple, rapid manipulations of complex material, in multiple forms, where term searching is important or the ability to overlay images or run animated simulations is involved, the technology is likely to offer many benefits, all else being equal, if the specific form is designed to be usable. Obviously, combining the technology with innovative classroom use, discretionary collaboration, and self-paced learning may offer further advantages, but as yet such scenarios remain largely unstudied.

Taking the literature as a whole, it is disappointing to report that statistical analyses and research methods are frequently flawed, limiting our understanding of these important issues. Failure to control important variables for comparative purposes, lack of adequate pretesting of learners, use of multiple *t* tests for post hoc data, and even the tendency to claim support for

hypotheses when the data fail to show statistically significant results all suggest that the basis for drawing conclusions from this literature is far from sturdy.

Coupled with this is a tendency for experimental investigations to limit themselves in terms of tasks, information types, and learners to very narrow areas of learning. Studies that involve as subjects individuals who are themselves learning to become instructors (McGrath, 1992) or that employ as test vehicles hypertexts about hypermedia (Jonassen & Wang, 1993) are self-referential and run the risk of failing to control for innumerable contaminating variables. We have no wish to prevent studies of instruction from being carried out on instructors; however, it is our hope that such work would be the exception, not the norm, for any conclusions we draw on learners and hypermedia.

How can this situation be improved? An immediate approach is to concentrate research efforts on those variables that seem most influential. Understanding the various components of learning tasks and identifying those that are likely to benefit directly from hypermedia interventions would be an obvious, human-factors focus that would surely reap benefits for our understanding (see Dillon, 1996, for details). The concept of "learning" is too broad to be equated with a simple statement about the pros and cons of any instructional medium, and we need to refine our analyses of tasks so that we can ensure greater generalizability of findings on constituent tasks such as information location, concept identification, reasoning, and categorization.

Thus, as indicated by the Psotka et al. (1993) study, careful task analysis of the learning requirement can highlight scenarios in which well-designed technology can facilitate the learning process. In this fashion, rather than claim a simple advantage or disadvantage for hypermedia in educational settings, efforts could be focused on those components of learning that are amenable to technological support so that maximum advantage of hypermedia's capabilities can be made. This is the manner in which one of the more successful exploitations of this technology was developed. Landauer et al. (1993) showed a gain for performance in an open-book statistics examination among users of their hypermedia application (SuperBook) after they iteratively designed and tested the system to remove usability problems.

A second focus is surely in the area of individual differences among learners. Although Chen and Rada (1996) concluded that differences in cognition alone do not have significant effects, they primarily examined active-passive cognitive-style studies and actually indicated that spatial ability is a medium-large effect. Since hypermedia may be implemented in multiple forms, those that are best suited to one style of learner may be completely inappropriate for others. The data reviewed here suggest that this is indeed the case, and a concerted research effort in this area is likely to offer considerable benefits.

Combining task analysis and learning differences in a strong program of empirical research would probably be the best approach. We remain convinced that well-designed hypermedia offers the potential to enhance learning

in a variety of ways. However, the design variables that are important are not well understood, and they interact with individual differences in learners that, at the moment, are not well understood. Identifying the precise combination of design attributes, task domains, and learner characteristics should be the focus of future research.

Beyond this pragmatic recommendation, there remains a tremendous need for a richer understanding of the learning process beyond how information presentation and access can enhance the educational experience. The use of this technology as a means of information creation by learners might be usefully explored, and there is a definite need to consider the potential for learning *with* hypermedia, not just from it. Jacobson (1994) made the case that cognitive flexibility theory offers potential insights into how this technology may best be used, and we do not disagree with his theory-to-design approach; however, we would make a call for stronger quantitative analysis of this approach (see e.g. Rouét, Levonen, Dillon, & Spiro, 1996). While the present review has the feel of a media comparison review, our aim has been more at informing the theoretical development in this field, since only by adequately accounting for findings of the kind reviewed here can a theory claim superiority over its competitors.

Acknowledgments

We would like to acknowledge the insightful comments of the anonymous reviewers and Carl Grant, who suggested changes, raised caveats, and caused us to focus the emphasis of this review considerably.

References

Aust, R., Kelley, M., & Roby, W. (1993). The use of hyper-reference and conventional dictionaries. *Educational Technology Research and Development*, *41*(4), 63–73.

Barrett, E. (1988). *Text, context and hypertext*. Cambridge, MA: MIT Press.

Becker, D., & Dwyer, M. (1994). Using hypermedia to provide learner control. *Journal of Educational Multimedia and Hypermedia*, *3*, 155–172.

Beishuizen, J., Stoutjesdijk, E., & van Putten, K. (1994). Studying textbooks: Effects of learning styles, study task, and instruction. *Learning and Instruction*, *4*, 151–174.

Blanchard, D. (1990). *A hypertext computer-assisted instruction (CAI) program: Teaching the Disk Operating System (DOS) to community college students*. Unpublished thesis, Center for the Advancement of Education, Nova University.

Chen, C., & Rada, R. (1996). Interacting with hypertext: A meta-analysis of experimental studies. *Human-Computer Interaction*, *11*, 125–156.

Collier, G. (1987). Thoth-II: Hypertext with explicit semantics. In *Proceedings of Hypertext '87* (pp. 269–289). Chapel Hill: University of North Carolina.

Delany, P., & Gilbert, S. (1991). Hypercard stacks for Fielding's *Joseph Andrews*: Issues of design and content. In P. Delaney & G. Landow (Eds.), *Hypermedia and literary studies* (pp. 287–298). Cambridge, MA: MIT Press.

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, *35*, 1297–1326.

Dillon, A. (1996). Myths, misconceptions and an alternative view of information usage and the electronic medium. In J. Rouét et al. (Eds.), *Hypertext and cognition* (pp. 25–42). Mahwah, NJ: Erlbaum.

Dillon, A., & Watson, C. (1996). User analysis HCI—The historical lessons from individual differences research. *International Journal of Human-Computer Studies*, *45*, 619–638.

Dryden, L. M. (1994). Literature, student-centered classrooms, and hypermedia environments. In C. Selfe & S. Hilligoss (Eds.), *Literacy and computers: The complications of teaching and learning with technology* (pp. 282–304). New York: Modern Language Association of America.

Eysenck, M. (1990). Personality and cognition. In M. Eysenck (Ed.), *Blackwell dictionary of cognitive psychology* (pp. 267–271). Oxford, England: Blackwell.

Frey, D., & Simonson, M. (1993). Assessment of cognitive style to examine students' use of hypermedia within historic costume. *Home Economics Research Journal*, *21*, 403–421.

Higgins, K., & Boone, R. (1990). Hypertext computer study guides and the social studies achievement of students with learning disabilities, remedial students, and regular education students. *Journal of Learning Disabilities*, 23, 529–540.

Hooper, S., & Hannafin, M. (1988, July). Learning the ROPES of instructional design: Guidelines for emerging interactive technologies. *Educational Technology*, pp. 14–17.

Horton, S., Boone, R., & Lovitt, T. (1990). Teaching social studies to learning disabled high school students: Effects of a hypertext study guide. *British Journal of Educational Technology*, *21*, 118–131.

Jacobson, M. (1994). Issues in hypertext and hypermedia research: Toward a framework for linking theory-to-design. *Journal of Educational Multimedia & Hypermedia*, *3(2)*, 141–154.

Jacobson, M., & Spiro, R. (1993). *Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation* (Tech. Rep. No. 573). Champaign: Center for the Study of Reading, University of Illinois.

Jonassen, D. (1989). *Hypertext/hypermedia*. Englewood Cliffs, NJ: Educational Technology Publications.

Jonassen, D., & Wang, S. (1993). Acquiring structural knowledge from semantically structured hypertext. *Journal of Computer-Based Instruction*, *20*, 1–8.

Kahn, P., Peters, R., & Landow, G. (1995). Three fundamental elements of visual rhetoric in hypertext. In W. Schuler, J. Hannemann, & N. Streitz (Eds.), *Designing user interfaces for hypermedia* (pp. 167–178). Berlin: Springer.

Landauer, T. (1995). *The trouble with computers*, Cambridge, MA: MIT Press.

Landauer, T., Egan, D., Remde, J., Lesk, M., Lochbaum, C., & Ketchum, D. (1993). Enhancing the usability of text through computer delivery and formative evaluation: The SuperBook Project. In C. McKnight, A. Dillon, & J. Richardson (Eds.), *Hypertext: A psychological perspective* (pp. 71–136). London: Ellis Horwood.

Landow, G. (1992). *Hypertext: The convergence of contemporary critical theory and technology*. Baltimore: Johns Hopkins University Press.

Landow, G., & Delany, P. (1991). Hypertext, hypermedia and literary studies: The state of the art. In P. Delaney & G. Landow (Eds.), *Hypermedia and literary studies* (pp. 3–50). Cambridge, MA: MIT Press.

Lanza, A., & Roselli, T. (1991). Effects of the hypertextual approach versus the structured approach on active and passive learners. *Journal of Computer-Based Instruction*, *18*(2), 48–50.

Lee, Y., & Lehman, J. (1993). Instructional cueing in hypermedia: A study with active and passive learners. *Journal of Educational Multimedia and Hypermedia*, *2*, 25–37.

Lehto, M., Zhu, W., & Carpenter, B. (1995). The relative effectiveness of hypertext and text. *International Journal of Human-Computer Interaction*, *7*, 293–313.

Lin, C., & Davidson, G. (1994, February). *Effects of linking structure and cognitive style on students' performance and attitude in a computer-based hypertext environment*. Paper presented at the annual convention of the Association for Educational Communications and Technology, Nashville, TN.

Liu, M., & Reed, M. (1994, February). *The relationship between the learning strategies and learning styles in a hypermedia environment*. Paper presented at the annual convention of the Association for Educational Communications and Technology, Nashville, TN.

Marchionini, G., & Crane, G. (1994). Evaluating hypermedia and learning: Methods and results from the Perseus Project. *ACM Transactions on Information Systems*, *12*, 5–34.

McGrath, D. (1992). Hypertext, CAI, paper or programs control: Do learners benefit from choices? *Journal of Research on Computing in Education*, *24*, 513–531.

McKnight, C., Dillon, A., & Richardson, J. (1991). *Hypertext in context*. Cambridge, England: Cambridge University Press.

McKnight, C., Dillon, A., & Richardson, J. (1996). User centered design of hypertext and hypermedia for education. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 622–633). New York: Macmillan.

McKnight, C., Dillon, A., Richardson, J., Haraldsson, H., & Spinks, R. (1992). Information access in different media: An experimental comparison. In E. J. Lovesey (Ed.), *Contemporary ergonomics 1992* (pp. 515–519). London: Taylor & Francis.

Nielsen, J. (1995). *Multimedia and hypertext: The Internet and beyond*. Cambridge, MA: Academic Press.

Psotka, J., Kerst, S., & Westerman, T. (1993). The use of hypertext and sensory-level supports for visual learning of aircraft names and shapes. *Behavior Research Methods*, *25*, 168–172.

Quade, A. (1993, January). *An assessment of the effectiveness of a hypertext instructional delivery system when compared to a traditional CAI tutorial*. Paper presented at the annual convention of the Association for Educational Communications and Technology, New Orleans, LA.

Recker, M., & Pirolli, P. (1995). Modeling individual differences in students' learning strategies. *Journal of the Learning Sciences*, *4*, 1–38.

Relan, A. (1991, January). The desktop environment in computer-based instruction: Cognitive foundations and implications for instructional design. *Educational Technology*, pp. 7–14.

Repman, J., Willer, H., & Lan, W. (1993). The impact of social context on learning in hypermedia-based instruction. *Journal of Educational Multimedia and Hypermedia*, *2*, 283–298.

Rouét, J., Levonen, J., Dillon, A., & Spiro, R. (Eds.) (1996). *Hypertext and cognition*. Mahaw, NJ: LEA.

Saga, H. (1992). Are we ready enough to learn from interactive multimedia? *Educational Media International*, *29*, 181–188.

Shin, E., Schallert, D., & Savenye, C. (1994). Effects of learner control, advisement, and prior knowledge on young students' learning in a hypertext environment. *Educational Technology Research and Development*, *42*, 33–46.

Shute, V. (1993). A comparison of learning environments: All that glitters. In: S. Lajoie & S. Derry (Eds.), *Computers as Cognitive Tools* (pp. 47–73). Hillsdale, NJ: LEA.

Smith, C. (1994). Hypertextual thinking. In C. Selfe & S. Hilligoss (Eds.), *Literacy and computers: The complications of teaching and learning with technology* (pp. 164–281). New York: Modern Language Association of America.

Stanton, N., & Baber, C. (1992). An investigation of styles and strategies in self-directed learning. *Journal of Educational Multimedia and Hypermedia*, *1*, 147–167.

Tripp, S., & Roby, W. (1990). Orientation and disorientation in a hypertext lexicon. *Journal of Computer-Based Instruction*, *17*(4), 120–124.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. London: Academic Press.

van der Berg, S. & Watt, J. (1991). Effects of educational setting on student responses to structured hypertext. *Journal of Computer-Based Instruction*, *18(4)*, 118–124.

Welsh, T., Murphy, K., Duffy, T., & Goodrum, D. (1993). Accessing elaborations on core information in a hypermedia environment. *Educational Technology Research & Development 41*(2), 19–34.

Part XVIII

COOPERATIVE GROUP WORK, PEER TUTORING

85

RESEARCH ON COOPERATIVE LEARNING AND ACHIEVEMENT

What we know, what we need to know

R. Slavin

Research on cooperative learning is one of the greatest success stories in the history of educational research. While there was some research on this topic from the early days of this century, the amount and quality of that research greatly accelerated in the early 1970s and continues unabated today, a quarter-century later. Hundreds of studies have compared cooperative learning to various control methods on a broad range of measures, but by far the most frequent objective of this research is to determine the effects of cooperative learning on student achievement. Studies of the achievement effects of cooperative learning have taken place in every major subject, at all grade levels, and in all types of schools in many countries. Both field studies and laboratory studies have produced a great deal of knowledge about the effects of many types of cooperative interventions and about the mechanisms responsible for these effects. Further, cooperative learning is not only a subject of research and theory, but it is also used at some level by millions of teachers. A recent national survey (Puma, Jones, Rock, & Fernandez, 1993) found that 79% of elementary teachers and 62% of middle school teachers reported making some sustained use of cooperative learning.

Given the substantial body of research on cooperative learning and the many cooperative learning programs in widespread use, it might be assumed that there is little further research to be done. Yet this is not the case. There are many very important questions in research on this topic, and a great deal of development and evaluation remains to be done. In its fullest conception, cooperative learning provides a radically different approach to instruction, whose possibilities have been tapped only on a limited basis.

While there is a growing consensus among researchers about the positive effects of cooperative learning on student achievement as well as a rapidly

Source: *Contemporary Educational Psychology*, 1996, 21, 43–69.

growing number of educators using cooperative learning at all levels of schooling and in many subject areas, there is still a great deal of confusion and disagreement about *why* cooperative learning methods affect achievement and, even more importantly, *under what conditions* cooperative learning has these effects. Researchers investigating cooperative learning effects on achievement have often operated in isolation from one another, almost on parallel tracks, and some describe theoretical mechanisms to explain achievement effects of cooperative learning that are totally different from the mechanisms assumed by others. In particular, there are researchers who emphasize the changes in incentive structure brought about by certain forms of cooperative learning, while others hold that changes in task structure are all that is required to enhance learning. The problem is that applications of cooperative learning typically change many aspects of both incentive and task structures, so disentangling which is responsible for which outcomes can be difficult.

In earlier writings, (Slavin, 1989, 1992, 1995), I identified four major theoretical perspectives (and two minor ones) designed to explain the achievement effects of cooperative learning. This paper updates and extends the discussion of these perspectives, further explores conditions under which each may operate, and suggests research and development needed to move the field of cooperative learning forward.

Four major theoretical perspectives on cooperative learning and achievement

Motivational perspectives

Motivational perspectives on cooperative learning focus primarily on the reward or goal structures under which students operate (see Slavin, 1977, 1983a, 1995). From a motivationalist perspective (e.g., Johnson & Johnson, 1992; Slavin, 1983a, 1983b, 1995), cooperative incentive structures create a situation in which the only way group members can attain their own personal goals is if the group is successful. Therefore, to meet their personal goals, group members must both help their groupmates to do whatever helps the group to succeed and, perhaps even more importantly, to encourage their groupmates to exert maximum efforts. In other words, rewarding groups based on group performance (or the sum of individual performances) creates an interpersonal reward structure in which group members will give or withhold social reinforcers (e.g., praise, encouragement) in response to groupmates' task-related efforts (see Slavin, 1983a). One intervention that uses cooperative goal structures is the group contingency (see Slavin, 1987), in which group rewards are given based on group members' behaviors. The theory underlying group contingencies does not require that group members be able to actually help one another or work together. The fact that their outcomes are dependent on one anothers' behavior is enough to motivate

students to engage in behaviors which help the group to be rewarded, because the group incentive induces students to encourage goal-directed behaviors among their groupmates (Slavin, 1983a, 1983b, 1995). A substantial literature in the behavior modification tradition has found that group contingencies can be very effective at improving students' appropriate behaviors and achievement (Hayes, 1976; Litow & Pumroy, 1975).

The motivationalist critique of traditional classroom organization holds that the competitive grading and informal reward system of the classroom creates peer norms opposing academic efforts (see Coleman, 1961). Since one student's success decreases the chances that others will succeed, students are likely to express norms that high achievement is for "nerds" or teachers' pets. Such work restriction norms are familiar in industry, where the "rate buster" is scorned by his or her fellow workers (Vroom, 1969). However, by having students work together toward a common goal, they may be motivated to express norms favoring academic achievement, to reinforce one another for academic efforts.

Not surprisingly, motivational theorists build group rewards into their cooperative learning methods. In methods developed by my colleagues and myself at Johns Hopkins University (Slavin, 1994, 1995), students can earn certificates or other recognition if their average team scores on quizzes or other individual assignments exceed a preestablished criterion (see also Kagan, 1992). Methods developed by Johnson and Johnson (1994) and their colleagues at the University of Minnesota often give students grades based on group performance, which is defined in several different ways. The theoretical rationale for these group rewards is that if students value the success of the group, they will encourage and help one another to achieve, much in contrast to the situation in the traditional, competitive classroom.

Empirical support for the motivational perspective

Evidence from practical applications of cooperative learning in elementary and secondary schools supports the motivationalist position that group rewards are essential to the effectiveness of cooperative learning, with one critical qualification. Use of group goals or group rewards enhances the achievement outcomes of cooperative learning if and only if the group rewards are based on the individual learning of all group members (Slavin, 1995). Most often, this means that team scores are computed based on average scores on quizzes which all teammates take individually, without teammate help. For example, in Student Teams-Achievement Divisions, or STAD (Slavin, 1994), students work in mixed-ability teams to master material initially presented by the teacher. Following this, students take individual quizzes on the material, and the teams may earn certificates based on the degree to which team members have improved over their own past records. The only way the team can succeed is to ensure that all team members have learned,

so the team members' activities focus on explaining concepts to one another, helping one another practice, and encouraging one another to achieve. In contrast, if group rewards are given based on a single group product (for example, the team completes one worksheet or solves one problem), there is little incentive for group members to explain concepts to one another, and one or two group members may do all the work (see Slavin, 1995).

A review of 99 studies of cooperative learning in elementary and secondary schools that involved durations of at least 4 weeks compared achievement gains in cooperative learning and control groups. Of 64 studies of cooperative learning methods that provided group rewards based on the sum of group members' individual learning, 50 (78%) found significantly positive effects on achievement, and none found negative effects (Slavin, 1995). The median effect size for the studies from which effect sizes could be computed was +.32 (32% of a standard deviation separated cooperative learning and control treatments). In contrast, studies of methods that used group goals based on a single group product or provided no group rewards found few positive effects, with a median effect size of only +.07. Comparisons of alternative treatments within the same studies found similar patterns; group goals based on the sum of individual learning performances were necessary to the instructional effectiveness of the cooperative learning models (e.g., Fantuzzo, Polite, & Grayson, 1990; Fantuzzo, Riggio, Connelly, & Dimeff, 1989; Huber, Bogatzki, & Winter, 1982). The importance of group goals and individual accountability is discussed further later in this paper.

Social cohesion perspectives

One theoretical perspective somewhat related to the motivational viewpoint holds that the effects of cooperative learning on achievement are strongly mediated by the cohesiveness of the group, in essence that students will help one another learn because they care about one another and want one another to succeed. This perspective is similar to the motivational perspective in that it emphasizes primarily motivational rather than cognitive explanations for the instructional effectiveness of cooperative learning. However, motivational theorists hold that students help their groupmates learn at least in part because it is in their own interests to do so. Social cohesion theorists, in contrast, emphasize the idea that students help their groupmates learn because they care about the group. A hallmark of the social cohesion perspective is an emphasis on teambuilding activities in preparation for cooperative learning and processing or group self-evaluation during and after group activities. Social cohesion theorists tend to downplay or reject the group incentives and individual accountability held by motivationalist researchers to be essential. For example, Cohen (1986, pp. 69–70) states "if the task is challenging and interesting, and if students are sufficiently prepared for skills in group process, students will experience the process of groupwork itself as highly

rewarding . . . never grade or evaluate students on their individual contributions to the group product." Cohen's (1994a) work, as well as that of Sharan and Sharan (1992) and Elliot Aronson (Aronson, Blaney, Stephan, Sikes, & Snapp, 1978) and their colleagues, may be described as social cohesion theories. Cohen, Aronson *et al.*, and Sharan and Sharan all use forms of cooperative learning in which students take on individual roles within the group, which Slavin (1983a) calls "task specialization" methods. In Aronson's Jigsaw method, students study material on one of four or five topics distributed among the group members. They meet in "expert groups" to share information on their topics with members of other teams who had the same topic, and then take turns presenting their topics to the team. In the Sharans' Group Investigation method, groups take on topics within a unit studied by the class as a whole and then further subdivide the topic into tasks within the group. The students investigate the topic together and ultimately present their findings to the class as a whole. Cohen's adaptation of DeAvila and Duncan's (1980) Finding Out/Descubrimiento program has students take different roles in discovery-oriented science activities.

One main purpose of the task specialization used in Jigsaw, Group Investigation, and Finding Out/Descubrimiento is to create interdependence among group members. In the Johnsons' methods, a somewhat similar form of interdependence is created by having students take on roles as "checker," "recorder," "observer," and so on. The idea is that if students value their groupmates (as a result of teambuilding and other cohesiveness-building activities) and are dependent on one another, they are likely to encourage and help one another to succeed. The Johnsons' (1989, 1994) work straddles the social cohesion and motivationalist perspectives described in this paper; while their models do use group goals and group incentives, their theoretical writings emphasize development of group cohesion through teambuilding, group self-evaluation, and other means more characteristic of social cohesion theorists.

Empirical support for the social cohesion perspective

The achievement outcomes of cooperative learning methods that emphasize task specialization are unclear. Research on the original form of Jigsaw has not generally found positive effects of this method on student achievement (Slavin, 1995). One problem with this method is that students have limited exposure to material other than that which they studied themselves, so learning gains on their own topics may be offset by losses on their groupmates' topics. In contrast, there is evidence that when it is well implemented, Group Investigation can significantly increase student achievement (Sharan & Shachar, 1988). In studies of at least 4 weeks' duration, the Johnsons' (1994) methods have not been found to increase achievement more than individualistic methods unless they incorporate group rewards (in this case, group

grades) based on the average of group members' individual quiz scores (see Slavin, 1995). Studies of forms of Jigsaw that have added group rewards to the original model have found positive achievement outcomes (Mattingly & Van Sickle, 1991).

Research on practical classroom applications of methods based on social cohesion theories provide inconsistent support for the proposition that building cohesiveness among students through teambuilding alone (i.e., without group incentives) will enhance student achievement. There is some evidence that group processing activities such as reflection at the end of each class period on the group's activities can enhance the achievement effects of cooperative learning (Yager, Johnson, Johnson, & Snider, 1986). On the other hand an Israeli study found that teambuilding activities had no effect on the achievement outcomes of Jigsaw (Rich, Amir, & Slavin, 1986).

In general, methods which emphasize teambuilding and group process but do not provide specific group rewards based on the learning of all group members are no more effective than traditional instruction in increasing achievement (Slavin, 1995), although there is evidence that these methods can be effective if group rewards are added to them. One major exception is Group Investigation (Sharan & Hertz-Lazarowitz, 1980; Sharan & Shachar, 1988; Sharan & Sharan, 1992). However, in this method groups are evaluated based on their group products, which are composed of unique contributions made by each group member. Thus, this method may be using a form of the group goals and individual accountability held by motivationalist theories to be essential to the instructional effectiveness of cooperative learning.

Cognitive perspectives

The major alternative to the motivationalist and social cohesiveness perspectives on cooperative learning, both of which focus primarily on group norms and interpersonal influence, is the cognitive perspective, which holds that interactions among students will in themselves increase student achievement for reasons which have to do with mental processing of information rather than with motivations. Cooperative methods developed by cognitive theorists involve neither the group goals that are the cornerstone of the motivationalist methods nor the emphasis on building group cohesiveness characteristic of the social cohesion methods. However, there are several quite different cognitive perspectives, as well as some which are similar in theoretical perspective but have developed on largely parallel tracks. These are described in the following sections.

Developmental perspectives

One widely researched set of cognitive theories is the developmental perspective (e.g., Damon, 1984; Murray, 1982). The fundamental assumption of the

developmental perspective on cooperative learning is that interaction among children around appropriate tasks increases their mastery of critical concepts. Vygotsky (1978, p. 86) defines the zone of proximal development as ". . . the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or *in collaboration with more capable peers*" (emphasis added). In his view, collaborative activity among children promotes growth because children of similar ages are likely to be operating within one another's proximal zones of development, modeling in the collaborative group behaviors more advanced than those they could perform as individuals. Vygotsky (1978) described the influence of collaborative activity on learning as follows:

> Functions are first formed in the collective in the form of relations among children and then become mental functions for the individual . . . Research shows that reflection is spawned from argument.

Similarly, Piaget (1926) held that social-arbitrary knowledge—language, values, rules, morality, and symbol systems—can only be learned in interactions with others. Peer interaction is also important in logical-mathematical thought in disequilibrating the child's egocentric conceptualizations and in provision of feedback to the child about the validity of logical constructions.

There is a great deal of empirical support for the idea that peer interaction can help non-conservers become conservers. Many studies have shown that when conservers and nonconservers of about the same age work collaboratively on tasks requiring conservation, the nonconservers generally develop and maintain conservation concepts (see Bell, Grossen, & Perret-Clermont, 1985; Murray, 1982; Perret-Clermont, 1980). In fact, a few studies (e.g. Ames and Murray, 1982; Mugny and Doise, 1978) have found that pairs of disagreeing nonconservers who had to come to consensus on conservation problems both gained in conservation. The importance of peers' operating in one anothers' proximal zones of development was demonstrated by Kuhn (1972), who found that a small difference in cognitive level between a child and a social model was more conducive to cognitive growth than a larger difference.

On the basis of these and other findings, many Piagetians (e.g., Damon, 1984; Murray, 1982; Wadsworth, 1984) have called for an increased use of cooperative activities in schools. They argue that interaction among students on learning tasks will lead *in itself* to improved student achievement. Students will learn from one another because in their discussions of the content, cognitive conflicts will arise, inadequate reasoning will be exposed, disequilibration will occur, and higher-quality understandings will emerge.

From the developmental perspective, the effects of cooperative learning on student achievement would be largely or entirely due to the use of cooperative

tasks. In this view, the opportunity for students to discuss, to argue, to present and hear one anothers' viewpoints is the critical element of cooperative learning with respect to student achievement. For example, Damon (1984, p. 335) integrates Piagetian, Vygotskian, and Sullivanian perspectives on peer collaboration to propose a "conceptual foundation for a peer-based plan of education:"

1. Through mutual feedback and debate, peers motivate one another to abandon misconceptions and search for better solutions.
2. The experience of peer communication can help a child master social processes, such as participation and argumentation, and cognitive processes, such as verification and criticism.
3. Collaboration between peers can provide a forum for discovery learning and can encourage creative thing.
4. Peer interaction can introduce children to the process of generating ideas.

However, Damon (1984, p. 337) explicitly rejects the use of "extrinsic incentives as part of the group learning situation," arguing that "there is no compelling reason to believe that such inducements are an important ingredient in peer learning."

One category of practical cooperative methods closely related to the developmental perspective is group discovery methods in mathematics, such as Marilyn Burns' (1981) Groups of Four method. In these techniques, students work in small groups to solve complex problems with relatively little teacher guidance. They are expected to discover mathematical principles by working with unit blocks, manipulatives, diagrams, and other concrete aids.

The theory underlying the presumed contribution of the group format is that in the exploration of opposing perceptions and ideas, higher-order understandings will emerge; also, students operating within one anothers' proximal zones of development will model higher-quality solutions for one another. However, studies of group discovery methods such as Groups of Four (Burns, 1981) find few achievement benefits for them in comparison to traditional expository teaching (Davidson, 1985; Johnson, 1985; Johnson & Waxman, 1985).

Despite considerable support from theoretical and laboratory research, there is little evidence from classroom experiments done over meaningful time periods that "pure" cooperative methods, which depend solely on interaction to produce higher achievement, will do so. However, it is likely that the cognitive processes described by developmental theorists are important as mediating variables to explain the effects of group goals and group tasks on student achievement (Slavin, 1987, 1995). This possibility is explored later in this paper.

Cognitive elaboration perspectives

A cognitive perspective on cooperative learning quite different from the developmental viewpoint is one which might be called the cognitive elaboration perspective. Research in cognitive psychology has long held that if information is to be retained in memory and related to information already in memory, the learner must engage in some sort of cognitive restructuring, or elaboration, of the material (Wittrock, 1986). One of the most effective means of elaboration is explaining the material to someone else. Research on peer tutoring has long found achievement benefits for the tutor as well as the tutee (Devin-Sheehan, Feldman, & Allen, 1976). Donald Dansereau and his colleagues at Texas Christian University have found in an impressive series of brief studies that college students working on structured "cooperative scripts" can learn technical material or procedures far better than can students working alone (Dansereau, 1988; O'Donnell, in press; Newbern, Dansereau, Patterson, & Wallace, 1994). In this method, students take roles as recaller and listener. They read a section of text, and then the recaller summarizes the information while the listener corrects any errors, fills in any omitted material, and helps think of ways both students can remember the main ideas. On the next section, the students switch roles. Dansereau and his colleagues found in a series of studies that while both the recaller and the listener learned more than did students working alone, the recaller learned more (O'Donnell & Dansereau, 1992). This mirrors both the peer tutoring findings and the findings of Noreen Webb (1989, 1992), who discovered that the students who gained the most from cooperative activities were those who provided elaborated explanations to others. In this research as well as in Dansereau's, students who received elaborated explanations learned more than those who worked alone, but not as much as those who served as explainers.

One practical use of the cognitive elaboration potential of cooperative learning is in writing process models (Graves, 1983), in which students work in peer response groups or form partnerships to help one another draft, revise, and edit compositions. Such models have been found to be effective in improving creative writing (Hillocks, 1984), and a writing process model emphasizing use of peer response groups is part of the Cooperative Integrated Reading and Composition Writing/Language Arts program (Stevens, Madden, Slavin, & Farnish, 1987), a program which has also been used to increase student writing achievement. Part of the theory behind the use of peer response groups is that if students learn to evaluate others' writing, they will become better writers themselves, a variant of the cognitive elaboration explanation. However, it is unclear at present how much of the effectiveness of writing process models can be ascribed to the use of cooperative peer response groups as opposed to other elements (such as the revision process itself).

One interesting development in recent years which relates to the cognitive elaboration perspective on cooperative learning is Reciprocal Teaching (Palincsar & Brown, 1984), a method for teaching reading comprehension skills. In this technique, students are taught to formulate questions for one another around narrative or expository texts. In doing so, they must process the material themselves and learn how to focus in on the essential elements of the reading passages. Studies of Reciprocal Teaching have generally supported its effects on student achievement (Palincsar, 1987; Rosenshine & Meister, 1994).

Reconciling the four perspectives

The four theoretical perspectives discussed above all have well-established rationales, and most have supporting evidence. All apply in some circumstances, but none are probably both necessary and sufficient in all circumstances. Research in each tradition tends to establish setting conditions favorable to that perspective. For example, most research on cooperative learning models from the motivational and social cohesiveness perspectives takes place in real classrooms over extended periods, as both extrinsic motivation and social cohesion may be assumed to take time to show their effects. In contrast, studies undertaken from the developmental and cognitive elaboration perspectives tends to be very short, making issues of motivation moot. These latter paradigms also tend to use pairs, rather than groups of four; pairs involve a much simpler social process than groups of four, which may need time to develop ways of working well together. Developmental research almost exclusively uses young children trying to master conservation tasks, which bear little resemblance to the "social-arbitrary" learning that characterizes most school subjects; cognitive elaboration research mostly involves college students.

However, the alternative perspectives on cooperative learning may be seen as complementary, not contradictory. For example, motivational theorists would not argue that the cognitive theories are unnecessary. Instead, they would argue that motivation drives cognitive process, which in turn produces learning. For example, it is unlikely that over the long haul students would engage in the kind of elaborated explanations found by Webb (1989) to be essential to profiting from cooperative activity. Similarly, motivational theorists would hold that an intermediate effect of extrinsic incentives must be to build cohesiveness, caring, and pro-social norms among group members, which could in turn affect cognitive processes. One model of the relationship among the four alternative perspectives is diagrammed in Fig. 1 (from Slavin, 1995).

The process depicted in Fig. 1 shows how group goals might operate to enhance the learning outcomes of cooperative learning. Provision of group goals based on the individual learning of all group members might affect

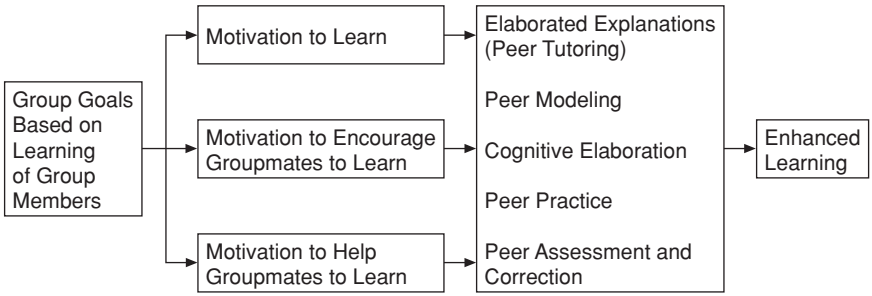


Figure 1

cognitive processes directly, by motivating students to engage in peer modeling, cognitive elaboration, and/or practice with one another. Group goals may also lead to group cohesiveness, increasing caring and concern among group members, making them feel responsible for one another's achievement, thereby motivating students to engage in cognitive processes which enhance learning. Finally, group goals may motivate students to take responsibility for one another independently of the teacher, thereby solving important classroom organization problems and providing increased opportunities for cognitively appropriate learning activities.

From the perspective of the model diagrammed in Fig. 1, researchers from outside of the motivational perspective are attempting to short-circuit the process to intervene directly on mechanisms identified as mediating variables in the full model. For example, social cohesion theorists intervene directly on group cohesiveness by engaging in elaborate teambuilding and group processing training. The Sharan and Shachar (1988) Group Investigation study suggests that this can be successfully done, but it takes a great deal of time and effort. In this study, teachers were trained over the course of a full year, and then teachers and students used cooperative learning for 3 months before the study began. Earlier research on Group Investigation failed to provide a comparable level of preparation of teachers and students, and the achievement results of these studies were less consistently positive (Sharan, *et al.*, 1984).

Cognitive theorists would hold that the cognitive processes that are essential to any theory relating cooperative learning to achievement can be created directly, without the motivational or affective changes discussed by the motivationalist and social cohesion theorists. This may turn out to be accurate, but at present demonstrations of learning effects from direct manipulation of peer cognitive interactions have mostly been limited to very brief durations and to tasks which lend themselves directly to the cognitive processes involved. For example, the Piagetian conservation tasks studied by developmentalists have few practical analogs in the school curriculum. However,

the research on Reciprocal Teaching in reading comprehension (Palincsar & Brown, 1984) shows promise as a means of intervening directly in peer cognitive processes. Long-term applications of Dansereau's (1988) cooperative scripts for comprehension of technical material and procedural instructions also seem likely to be successful.

What factors contribute to achievement effects of cooperative learning?[1]

Research on cooperative learning has moved beyond the question of whether cooperative learning is effective in accelerating student achievement to focus on the conditions under which it is optimally effective. The foregoing discussion describes alternative overarching theories to explain cooperative learning effects, and an integration of these theories. Beyond this, it is important to understand in more detail the factors that contribute to or detract from the effectiveness of cooperative learning.

There are two primary ways to learn about factors that contribute to the effectiveness of cooperative learning. One is to compare the outcomes of studies of alternative methods. For example, if programs that incorporated group rewards produced stronger or more consistent positive effects (in comparison to control groups) than programs that did not, then this would provide one kind of evidence that group rewards enhance the outcomes of cooperative learning. The problem with such comparisons is that the studies being compared usually differ in measures, durations, subjects, and many other factors that could explain differing outcomes. Better evidence is provided by studies that compared alternative forms of cooperative learning. In such studies, most factors, other than the ones being studied, can be held constant. The following sections discuss both types of studies to further explore factors that contribute to the effectiveness of cooperative learning for increasing achievement.

Group goals and individual accountability

As noted earlier, reviewers of the cooperative learning literature have long concluded that cooperative learning has its greatest effects on student learning when groups are recognized or rewarded based on individual learning of their members (Slavin, 1983a, 1983b, 1989, 1992, 1995; Ellis & Fouts, 1993; Newmann & Thompson, 1987; Manning & Lucking, 1991; Davidson, 1985; Mergendoller & Packer, 1989). For example, methods of this type may give groups certificates based on the average of individual quiz scores of group members, where group members could not help each other on the quizzes. Alternatively, group members might be chosen at random to represent the group, and the whole group might be rewarded based on the selected member's performance. In contrast, methods lacking group goals give students

only individual grades or other individual feedback, and there is no group consequence for doing well as a group. Methods lacking individual accountability might reward groups for doing well, but the basis for this reward would be a single project, worksheet, quiz, or other product that could theoretically have been done by only one group member.

The importance of group goals and individual accountability is in providing students with an incentive to help each other and to encourage each other to put forth maximum effort (Slavin, 1995). If students value doing well as a group, and the group can succeed only by ensuring that all group members have learned the material, then group members will be motivated to teach each other. Studies of behaviors within groups that relate most to achievement gains consistently show that students who give each other explanations (and less consistently, those who receive such explanations) are the students who learn the most in cooperative learning. Giving or receiving answers without explanation generally reduces achievement (Webb, 1989, 1992). At least in theory, group goals and individual accountability should motivate students to engage in the behaviors that increase achievement and avoid those that reduce it. If a group member wants her group to be successful, she must teach her groupmates (and learn the material herself). If she simply tells her groupmates the answers, they will fail the quiz that they must take individually. If she ignores a groupmate who is not understanding the material, the groupmate will fail and the group will fail as well. In groups lacking individual accountability, one or two students may do the group's work, while others engage in "social loafing" (Latane, Williams, & Harkins, 1979). For example, in a group asked to complete a single project or solve a single problem, some students may be discouraged from participating. A group trying to complete a common problem may not want to stop and explain what is going on to a groupmate who doesn't understand, or may feel it is useless or counterproductive to try to involve certain groupmates.

The evidence from research on cooperative learning strongly supports the importance of group goals that can be achieved only by ensuring the learning of all group members. The most recent comprehensive review of this topic by Slavin (1995) provides one kind of evidence to support this conclusion. Studies of methods that incorporated group goals and individual accountability produced a much higher median effect size than did studies of other methods. As noted earlier, the median effect size across 52 studies was +.32, compared to a median of only +.07 across 25 studies that did not incorporate group goals and individual accountability. Seventy-eight percent of studies of methods using group goals and individual accountability found significantly positive effects, and there were no significantly negative effects. In methods lacking these elements only 37% of studies found significantly positive effects, and 14% found significantly negative effects.

A comparison among Learning Together studies (Johnson & Johnson, 1989) also supports the same conclusions. Across eight studies of Learning

Together methods in which students were rewarded based on a single worksheet or product, the median effect size was near zero (+.04). However, among four studies that evaluated forms of the program in which students were graded based on the average performance of all group members on individual assessments, three found significantly positive effects.

Finally, comparisons within the same studies consistently support the importance of group goals and individual accountability. For example, Fantuzzo, King, and Heller (1992) conducted a component analysis of Reciprocal Peer Tutoring (RPT). They compared four conditions in which students worked in dyads to learn math. In one, students were rewarded with opportunities to engage in special activities of their choice if the sum of the dyad's scores on daily quizzes exceeded a criterion. In another, students were taught a structured method of tutoring each other, correcting efforts, and alternating tutor–tutee roles. A third condition involved a combination of rewards and structure, and a fourth was a control condition in which students worked in pairs but were given neither rewards nor structure. The results showed that the reward + structure condition had by far the largest effects on math achievement (ES = +1.42), and that reward alone had much larger effects than structure alone. The reward + structure condition exceeded structure-only by an effect size of +1.88, and the reward-only group exceeded control by an effect size of +.21 (the structure-only group performed less well than the control group).

Other studies also found greater achievement for cooperative methods using group goals and individual accountability than for those that do not. Huber, Bogatzki, and Winter (1982) compared a form of STAD to traditional group work lacking group goals and individual accountability. The STAD group scored significantly better on a math test (ES = +.23). In a study of TAI, Cavanaugh (1984) found that students who received group recognition based on the number of units accurately completed by all group members both learned more (ES = +.24) and completed more units (ES = +.25) than did students who received individual recognition only. O'Donnell (in press) compared dyads working with and without incentives. Students who received explicit incentives based on their learning learned significantly more than those who did not in three experimental studies. Okebukola (1985), studying science in Nigeria, found substantially greater achievement in STAD and TGT, methods using group goals and individual accountability, than in forms of Jigsaw and Johnsons' methods that did not. In another study, Okebukola found much higher achievement in classes that used a method combining cooperation and group competition (one form of group reward) than in a "pure" cooperative method that did not use group rewards of any kind (ES = +1.28).

A few reviewers (e.g., Damon, 1984; Kohn, 1986) have recommended against the use of group rewards, fearing that they may undermine long-term motivation. There is no evidence that they do so, and they certainly do

not undermine long-term achievement. Among multi-year studies, methods that incorporate group rewards based on individual learning performance have consistently shown continued or enhanced achievement gains over time (Stevens & Slavin, 1995a, 1995b; Hertz-Lazarowitz, Ivory, & Calderón, 1993; Greenwood, Delquadri, & Hall, 1989). In contrast, multi-year studies of methods lacking group rewards found few achievement effects in the short or long term (Solomon, Watson, Schaps, Battistich, & Solomon, 1990; Talmage, Pascarella, & Ford, 1984).

Cohen (1994b) raises the possibility that while group rewards and individual accountability may be necessary for lower-level skills, they may not be for higher-level ones. As evidence of this she cites a study by Sharan *et al.* (1984) that compared STAD and Group Investigation. In this study STAD and GI students performed equally well (and better than controls) on a test of English as a foreign language, and STAD students did significantly better than GI on "lower level" (knowledge) items (ES = +.38). On "high level" items, GI students performed nonsignificantly higher than STAD students, with a difference of less than half of a point on a 15-point test. Otherwise there is no evidence that group rewards are less important for higher-order skills, although the possibility is intriguing.

Structuring group interactions

While it is clear that all other things being equal, group rewards and individual accountability greatly enhance the achievement outcomes of cooperative learning, there is some evidence that carefully structuring the interactions among students in cooperative groups can also be effective, even in the absence of group rewards. For example, Meloth and Deering (1992) compared students working in two cooperative conditions. In one, students were taught specific reading comprehension strategies and given "think sheets" to remind them to use these strategies (e.g., prediction, summarization, character mapping). In the other group, students earned team scores if their members improved each week on quizzes. A comparison of the two groups on a reading comprehension test found greater gains for the strategy group (also see Meloth & Deering, 1994). Berg (1993) and Newbern *et al.* (1994) found positive effects of scripted dyadic methods that did not use group rewards, and Van Oudenhoven, Wiersma, and Van Yperen (1987) found positive effects of structured pair learning whether feedback was given to the pairs or only to individuals.

Research on Reciprocal Teaching (Palincsar & Brown, 1984) also shows how direct strategy instruction can enhance the effects of a technique related to cooperative learning. In this method, the teacher works with small groups of students and models such cognitive strategies as question generation and summarization. The teacher then gradually turns over responsibility to the students to carry on these activities with each other. Studies of Reciprocal

Teaching have generally found positive effects of this method of reading comprehension (Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987; Rosenshine & Meister, 1994). The effects of group rewards based on the individual learning of all group members are clearly indirect; they only motivate students to engage in certain behaviors, such as providing each other with elaborated explanation. The research by Meloth and Deering (1992), Berg (1993), and others suggests that students can be directly taught to engage in cognitive and interpersonal behaviors that lead to higher achievement, without the need for group rewards.

However, there is also a growing body of evidence to suggest that a combination of group rewards and strategy training produces much better outcomes than either alone. First, the Fantuzzo *et al*. (1992) study, cited earlier, directly made a comparison between rewards alone, strategy alone, and a combination, and found the combination to be by far the most effective. Further, the outcomes of the RPT and CWPT dyadic learning methods, which use group rewards as well as strategy instruction, produced some of the largest positive effects of any cooperative methods, much larger than those found in the Berg (1993) study that provided groups with structure but not rewards. As noted earlier, studies of scripted dyads also find that adding incentives adds to the effects of these strategies (O'Donnell, in press). The consistent positive findings for CIRC, which uses both group rewards and strategy instruction, also argue for this combination.

Which students gain most from cooperative learning?

Several studies have focused on the question of which students gain the most from cooperative learning. One particularly important question relates to whether cooperative learning is beneficial to students at all levels of prior achievement. It would be possible to argue (see, for example, Allan, 1991; Robinson, 1990) that high achievers could be held back by having to explain material to their low-achieving groupmates. However, it would be equally possible to argue that because students who give elaborated explanations typically learn more than those who receive them (Webb, 1992), high achievers should be the students who benefit most from cooperative learning because they give the most frequent elaborated explanations.

The evidence from experimental studies that met the inclusion criteria for this review support neither position. A few studies found better outcomes for high achievers than for low and a few found that low achievers gained the most (see Slavin, 1995). Most, however, found equal benefits for high, average, and low achievers in comparison to their counterparts in control groups. One 2-year study of schools using cooperative learning most of their instructional day found that high, average, and low achievers all achieved better than controls at similar achievement levels. However, a separate analysis of the very highest achievers, those in the top 10% and top 5% of their

classes at pretest, found particularly large positive effects of cooperative learning on these students (Slavin, 1991; Stevens & Slavin, 1995b).

A few studies have looked for possible differences in the effects of cooperative learning on students of different ethnicities. Several have found particularly large effects for black students. However, other studies have found equal effects of cooperative learning for students of different backgrounds (Slavin, 1995).

Other studies have examined a variety of factors that might interact with achievement gain in cooperative learning. Okebukola (1986) and Wheeler and Ryan (1973) found that students who preferred cooperative learning learned more in cooperative methods than those who preferred competition. Chambers and Abrami (1991) found that students on successful teams learned more than those on less successful teams.

Finally, a small number of studies have compared variations in cooperative procedures. Moody and Gifford (1990) found that while there was no difference in achievement gains of homogeneous and heterogeneous groups, pairs performed better than groups of four and gender-homogeneous groups performed better than mixed groups. Foyle, Lyman, Tompkins, Perne, and Foyle (1993) found that cooperative learning classes assigned daily homework achieved more than those not assigned homework. Kaminski (1991) and Rich *et al*. (1986) found that explicit teaching of collaborative skills had no effect on student achievement. Jones (1990) compared cooperative learning using group competition to an otherwise identical method that compared groups to a set standard (as in STAD). There were no achievement differences, but a few attitude differences favored the group competition.

Are group goals and individual accountability always necessary?

The previous discussion has summarized evidence that generally supports the motivationalist view that group goals and individual accountability are necessary for cooperative learning to result in achievement gains, at least in applications of several weeks or months (my 1995 review considered only studies of at least four weeks' duration). Yet there are a few cases in which achievement gains (in comparison to control groups) have been found for cooperative learning treatments that lack one or both of these elements. Are there conditions under which they may not be necessary?

Before exploring this question, it is important to further consider the theoretical rationale for the importance of group goals and individual accountability. Both are principally designed to motivate students to teach each other, to be concerned about the learning of their groupmates. The assumption behind them is that while groupmates may readily interact with each other and help each other, without appropriate structuring this interaction and help may take the form of sharing answers or doing each other's work rather than making certain that groupmates can independently solve problems or

know the material. In cooperative learning techniques in which groups are rewarded based on the individual learning of each group member, the group members want the group to succeed, and the only way they can make this happen is to teach and assess one another to make certain that every group member can independently show mastery of whatever the group is studying.

The theoretical and empirical support for the centrality of group goals and individual accountability is strong for a broad range of school tasks. Yet there may be some kinds of tasks that do not require these elements.

Controversial tasks without single answers

One category of tasks that may not require group goals and individual accountability is tasks in which it is likely that students will benefit by hearing others thinking aloud. This is the classic Vygotskian paradigm; students in collaborating groups make overt their private speech, giving peers operating at a slightly lower cognitive level on a given task a stepping stone to understanding and incorporating higher-quality solutions in their own private speech (see Bershon, 1992). Tasks of this kind would be ones at a very high level of cognitive complexity but without a well-defined path to a solution or a single correct answer, especially tasks on which there are likely to be differences of opinion. For such tasks, the process of participating in arguments or even of listening to others argue and justify their opinions or solutions may be enough to enhance learning, even if no teaching, explanation, or assessment goes on within the group. Perhaps the best classroom evidence on this type of task is from Johnson and Johnson's (1979) studies of structured controversy, in which students argue both sides of a controversial issue using a structured method of argumentation. Other examples of such tasks might include group projects without a single right answer (e.g., planning a city), and solving complex problems, such as nonroutine problems in mathematics or finding the main idea of paragraphs. In each of these cases it may be that hearing others' thinking processes is beneficial even if co-teaching does not take place.

It is still important to note that use of group goals and individual accountability is unlikely to interfere with modeling of higher-level thinking and is likely to add teaching and elaborated explanation (Webb, 1992). For example, Stevens, Slavin, and Farnish (1991) evaluated a method of teaching students to find the main ideas of paragraphs in which four-member groups first came to consensus on a set of paragraphs and then worked to make certain that every group member could find the main idea. Groups received certificates based on the performance of their members on individual quizzes. The consensus procedure evoked arguments and explanations, modeling higher quality thinking, but the teaching procedure made sure that students could each apply their new understandings. The program's effects on tests of main idea and inference skills were substantial.

Voluntary study groups

A second category of cooperative tasks that may not require group goals and individual accountability is situations in which students are strongly motivated to perform well on an external assessment and can clearly see the benefit of working together. The classic instance of this is voluntary study groups common in postsecondary education, especially in medical and law schools. Medical and law students must master an enormous common body of information, and it is obvious to many students that participating in a study group will be beneficial. While there is little extrinsic reason for students to be concerned about the success of other study group members, there is typically a norm within study groups that each member must do a good job of presenting to the group. Because study group membership is typically voluntary, study group members who do not participate effectively may be concerned that next term they may not be invited back.

There is little research on voluntary study groups in postsecondary institutions, and it is unclear how well this idea would apply at the elementary or secondary levels. In the United States, it would seem that only college-bound senior high school students are likely to care enough about their grades to actively participate in study groups like those seen at the postsecondary level, yet it may be that similar structures could be set up by teachers and that norms of reciprocal responsibility to the group could be developed. Another problem, however, is that voluntary study groups can and do reject (or fail to select) members who are felt to have little to contribute to the group. This could not be allowed to happen in study groups sponsored by the school.

Structured dyadic tasks

A third category of cooperative tasks that may not require group goals and individual accountability is tasks that are so structured that learning is likely to result if students engage in them, regardless of their motivation to help their partners learn. Examples of this were discussed earlier. One is the series of studies by Dansereau (1988) and his colleagues in which pairs of college students proceeded through a structured sequence of activities to help each other learn complex technical information or procedures (see O'Donnel & Dansereau, 1992). Another was two Dutch studies of spelling which also involved dyads and in which the study behavior (quizzing each other in turn) was structured and obviously beneficial (Van Oudenhoven, Wiersma, & Van Yperen, 1987; Van Oudenhoven, Van Berkum, & Koopmans, 1987). In contrast to cooperative methods using group goals and individual accountability to indirectly motivate students to teach each other, these methods allow the teacher to directly motivate students to engage in structured turn-taking behaviors known to increase learning. The successful use of structured

dyadic tasks in elementary schools seems largely limited to lower-level, rote skills such as memorizing multiplication tables, spelling lists, or place names.

As in the case of controversial tasks without single correct answers, there is evidence that adding group rewards to structured dyadic tasks enhances the effects of these strategies. Fantuzzo, Polite, and Grayson (1990) evaluated a dyadic study strategy called Reciprocal Peer Tutoring. A simple pair study format did not increase student arithmetic achievement, but when successful dyads were awarded stickers and classroom privileges, their achievement markedly increased. A similar comparison of dyadic tutoring with and without group rewards at the college level also found that group rewards greatly enhanced the achievement effects of a structured dyadic study model (Fantuzzo, Riggio, Connelly, & Dimeff, 1989), and a series of studies have shown positive effects of the Reciprocal Peer Tutoring model in many subjects and at many grade levels (e.g., Fantuzzo, Polite, & Grayson, 1990). A similar program combining structured reciprocal tutoring with group rewards called Classwide Peer Tutoring has also been successful in increasing student achievement in a variety of subjects and grade levels (Greenwood, Delquardi, & Hall, 1989; Maheady, Harper, & Mallette, 1991).

Needs for additional research

The four theoretical models explaining the achievement effects of cooperative learning described in this paper are all useful in expanding our understanding of the conditions under which various forms of cooperative learning may affect student achievement. Figure 1, which links these theoretical perspectives in a causal model, provides a framework for predicting different causal paths by which cooperative learning might affect achievement.

In particular, the model shows the importance of group goals and individual accountability, but also suggests ways that achievement might be affected more directly by introducing peer activities that may not require extrinsic motivation. This paper explores three types of tasks or situations in which group goals and individual accountability may not be necessary: controversial tasks lacking single right answers, voluntary study groups, and structured dyadic tasks. There is no research on voluntary study groups (such as medical or law school study groups), but research does find instances in which the other two types of cooperative tasks are effective without group goals and individual accountability. However, there is also evidence that adding group goals and individual accountability to these tasks further enhances their instructional effectiveness.

Clearly, there is a need for further research on conditions under which group goals and individual accountability may not be necessary. As a practical matter, it is probably the case that most teachers using cooperative learning do not provide group rewards based on the individual learning of

all group members, and feel that it is unnecessary and cumbersome to do so. Widespread reluctance to use extrinsic incentives, based in part on a misreading of research on the 'undermining' effects of rewards on long-term motivation (Cameron & Pierce, 1994) has contributed to many educators' reluctance to use group rewards. For both theoretical and practical reasons it would be important to know how to make 'reward-free' cooperative learning methods effective.

A related need for research concerns effective uses of project-based learning. Most research on cooperative learning has involved the use of these methods to help children master fairly well-defined skills or information. They key exceptions to this are work of the Sharan and Sharan (1992) and of Elizabeth Cohen (1994b). Cooperative learning practice has increasingly shifted toward project-based or active learning (Stern, in press), in which students work together to produce reports, projects, experiments, and so on. It is possible to make inferences to optimal conditions for project-based learning from research on more cut-and-dried content (see Slavin, in press), and the work of Cohen and the Sharans does imply that well-implemented, project-based learning can be more effective than traditional instruction (the Sharan & Shachar [1988] study is by far the best evidence of this). However, there is a great deal of work yet to be done to identify effective, replicable methods, to understand the conditions necessary for success in project-based learning, and to develop a more powerful theory and rationale to support project-based learning.

There is a need for both development and research at the intersection of cooperative learning and curriculum. Our own work has for many years focused on development and evaluation of cooperative learning methods that are tied to particular subjects and grade levels, such as Cooperative Integrated Reading and Composition (Stevens *et al.*, 1987) and our newest programs, WorldLab (social studies and science) and Math Wings (Slavin, Madden, Dolan, & Wasik, 1994). Elizabeth Cohen's (1994a) Complex Instruction program and Eric Schaps' (Soloman *et al.*, 1990) Child Development Project have also developed specific, broadly applicable curriculum materials to be used in a cooperative learning format. These contrast with most cooperative learning models which typically provide some general guidance for how to adapt cooperative learning to different subjects and grade levels but rarely provide actual student materials. How is cooperative learning affected by the existence of specific materials? Does use of these materials improve the learning outcomes of cooperative learning? Does it make cooperative learning more likely to be implemented well in the first place and maintained over time? Or does the use of prepared materials lead to less thoughtful use of cooperative learning or less ability to adapt in situations lacking materials? These questions are more important for practice than for theory but they are very important for practice. Not incidentally, there is a need for development of high-quality well-developed, well-researched

cooperative curricula in many subjects and grade levels, especially at the secondary level.

Related to the need for research on curriculum-based methods is a need for research on effective strategies for professional development and follow up to support cooperative learning. Nearly all cooperative learning training programs make extensive use of simulations, and this seems so obviously effective that perhaps it is not worth studying (although perhaps it is at least worth documenting). There has been some research on the effectiveness of peer coaching to support implementations of cooperative learning (e.g., Joyce, Hersh, & McKibbin, 1983). Yet there is much more work to be done to identify strategies for professional development likely to lead to high-quality, thoughtful, and sustained implementation. A few factors worth studying might include contrasts between schoolwide and teacher-by-teacher implementations, expert versus peer coaches, inservice focusing on generic principles versus specific strategies, and use of teacher learning communities (Calderón, 1994), groups of teachers who meet on a regular basis to support each other's innovative efforts.

Perhaps the only determined opposition to cooperative learning within the community of professional educators has come from advocates for gifted students. There is some research on the effects of cooperative learning on gifted students both within heterogeneous classes (Stevens & Slavin, 1995b) and within separate programs for the gifted (Gallagher, 1995), and so far there is little evidence to support fears that gifted students are shortchanged by cooperative learning. However, much more research is needed in this area to understand, for example, whether different cooperative methods have different effects for gifted students and how the effects of cooperative learning might be different in homogeneous and heterogeneous settings. On this last question, there is a broader need to study cooperative learning in the context of attempts to replace homogeneous with heterogenous grouping, especially in middle and high schools, and to use cooperative learning instead of homogenous reading groups in elementary schools.

This paper has focused on the achievement outcomes of cooperative learning, but there are of course many other outcomes that are in need further research. Among there are studies of cooperative learning effects on intergroup relations, self-esteem, acceptance of mainstreamed classmates, pro-social norms, and so on (see Slavin, 1995; Hawley & Jackson, 1995).

In general, there is a need for more research on all outcomes in senior high schools and in post-secondary institutions and a need for development and evaluations of cooperative methods for young children, especially those in pre-kindergarten, kindergarten, and first grade.

In summary, although cooperative learning has been studied in an extraordinary number of field experiments of high methodological quality, there is still much more to be done. Cooperative learning has the potential to become a primary format used by the teachers to achieve both traditional

and innovative goals. Research must continue to provide the practical, theoretical, and intellectual underpinnings to enable educators to achieve this potential.

Acknowledgments

This paper is adapted from Slavin (1992). It was written under a grant from the Office of Educational Research and Improvement, U.S. Department of Education (OERI-R-117-D40005). However, any opinions expressed are mine and do not necessarily represent OERI positions or policies.

Note

1 These sections are adapted from Slavin (1995).

References

Allen, S. D. (1991). Ability grouping research reviews: What do they say about grouping and the gifted? *Educational Leadership*, **48**(6), 60–65.

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive change by social conflict. *Developmental Psychology*, **18**, 894–897.

Aronson, E., Blaney, N., Stephan, C., Sikes, J., & Snapp, M. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage.

Bell, N., Grossen, M., & Perret-Clermont, A-N. (1985). Socio-cognitive conflict and intellectual growth. In M. Berkowitz (Ed.), *Peer conflict and psychological growth*. San Francisco: Jossey-Bass.

Berg, K. F. (1993, April). *Structured cooperative learning and achievement in a high school mathematics class*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Bershon, B. L. (1992). Cooperative problem solving: A link to inner speech. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups* (pp. 36–48). New York: Cambridge Univ. Press.

Burns, M. (1981, September). Groups of four: Solving the management problem. *Learning*, pp. 46–51.

Calderón, M. (1994). Mentoring and coaching minority teachers. In J. DeVillar & J. Cummins (Eds.), *Successful cultural diversity: Classroom practices for the 21st century*. Albany, NY: SUNY Press.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, **64**, 363–423.

Cavanagh, B. R. (1984). Effects of interdependent group contingencies on the achievement of elementary school children. *Dissertation Abstracts*, **46**, 1558.

Chambers, B., & Abrami, P. C. (1991). The relationship between Student Team Learning outcomes and achievement, causal attributions, and affect. *Journal of Educational Psychology*, **83**, 140–146.

Cohen, E. (1986). *Designing groupwork: Strategies for the heterogeneous classroom*. New York: Teachers College Press.

Cohen, E. G. (1994a). *Designing groupwork: Strategies for the heterogeneous classroom* (2nd ed.). New York: Teachers College Press.

Cohen, E. G. (1994b). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, **64**(1), 1–35.

Coleman, J. (1961). *The adolescent society*. New York: Free Press.

Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology*, **5**, 331–343.

Dansereau, D. F. (1988). Cooperative learning strategies. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 103–120). Orlando, FL: Academic Press.

Davidson, N. (1985). Small–group learning and teaching in mathematics: A selective review of the research. In R. E. Slavin, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb, & R. Schmuck (Eds.). *Learning to cooperating to learn* (pp. 211–230). New York: Plenum.

De Avila, E., & Duncan, S. (1980). *Finding out/descubrimiento*. Corte Madera, CA: Linguametrics Group.

Devin-Sheehan, L., Feldman, R., & Allen, V. (1976). Research on children tutoring children: A critical review. *Review of Educational Research*, **46**(3), 355–385.

Ellis, A. K., & Fouts, J. T. (1993). *Research on educational innovations*. Princeton Junction, NJ: Eye on Education.

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology*, **84**, 33–339.

Fantuzzo, J. W., Polite, K., & Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *Journal of School Psychology*, **28**, 309–323.

Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, **81**, 173–177.

Foyle, H. C., Lyman, L. R., Tompkins, L., Perne, S., & Foyle, D. (1993). Homework and cooperative learning: A classroom field experiment. *Illinois School Research and Development*, **29**(3), 25–27.

Gallagher, J. J. (1995). Educational of gifted students: A civil rights issue? *Phi Delta Kappan*, **76**(5), 408–410.

Graves, D. (1983). *Writing: Teachers and children at work*. Exeter, NH: Heinemann.

Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, **81**, 371–383.

Hawley, W. D., & Jackson, A. W. (Eds.). (1995). *Toward a common destiny: Improving race and ethnic relations in America*. San Francisco: Jossey-Bass.

Hayes, L. (1976). The use of group contingencies for behavioral control: A review. *Psychological Bulletin*, **83**, 628–648.

Hertz-Lazarowitz, R., Ivory, G., & Calderón, M. (1993). *The Bilingual Cooperative Integrated Reading and Composition (BCIRC) project in the Ysleta Independent School District: Standardized test outcomes*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Hillocks, G. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. *American Journal of Education*, **93**, 133–170.

Huber, G. L., Bogatzki, W., & Winter, M. (1982). *Kooperation als Ziel schulischen Lehrens und Lehrens*. Tubingen, West Germany: Arbeitsbereich Padagogische Psychologie der Universitat Tubingen.

Johnson, D. W., & Johnson, R. T. (1979). Conflict in the classroom: Controversy and learning. *Review of Educational Research*, **49**, 51–70.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book Co.

Johnson, D. W., & Johnson, R. T. (1992). Positive interdependence: Key to effective cooperation. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 174–199). New York: Cambridge Univ. Press.

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive, and individualistic learning* (4th ed.). Boston: Allyn & Bacon.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, **89**, 47–62.

Johnson, L. C. (1985). *The effects of the groups of four cooperative learning model on student problem-solving achievement in mathematics*. Unpublished doctoral dissertation, University of Houston.

Johnson, L. C., & Waxman, H. C. (1985, March). *Evaluating the effects of the "groups of four" program*. Paper presented at the annual convention of the American Educational Research Association, Chicago.

Jones, D. S. P. (1990). *The effects of contingency based and competitive reward systems on achievement and attitudes in cooperative learning situations*. Unpublished doctoral dissertation, Temple University.

Joyce, B. R., Hersh, R. H., & McKibbin, M. (1983). *The structure of school improvement*. New York: Longman.

Kagan, S. (1992). *Cooperative learning* (8th ed.). San Juan Capistrano, CA: Kagan Cooperative Learning.

Kaminski, L. B. (1991). *The effect of formal group skill instruction and role development on achievement of high school students taught with cooperative learning*. Unpublished doctoral dissertation, Michigan State University.

Kohn, A. (1986). *No contest: The case against competition*. Boston: Houghton-Mifflin.

Kuhn, D. (1972). Mechanism of change in the development of cognitive structures. *Child Development*, **43**, 833–844.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, **37**, 822–832.

Litow, L., & Pumroy, D. (1975). A brief review of classroom group-oriented contingencies. *Journal of Applied Behavior Analysis*, **8**, 341–347.

Maheady, L., Harper, G. F., & Mallette, B. (1991). Peer-mediated instruction: Review of potential applications for special education. *Reading, Writing, and Learning Disabilities*, **7**, 75–102.

Mandl, H., & Renkl, A. (1993). A plea for "more local" theories of cooperative learning. *Learning and Instruction*, **3**, 12–18.

Manning, M. L., & Lucking, R. (1991, May/June). The what, why, and how of cooperative learning. *The Social Studies*, pp. 120–124.

Mattingly, R. M., & Van Sickle, R. L. (1991). Cooperative learning and achievement in social studies: Jigsaw II. *Social Education*, **55**(6), 392–395.

Meloth, M. S., & Deering, P. D. (1992). The effects of two cooperative conditions on peer group discussions, reading comprehension, and metacognition. *Contemporary Educational Psychology*, **17**, 175–193.

Meloth, M. S., & Deering, P. D. (1994). Task talk and task awareness under different cooperative learning conditions. *American Educational Research Journal*, **31**(1), 138–166.

Mergendoller, J., & Packer, M. J. (1989). *Cooperative learning in the classroom: A knowledge brief on effective teaching*. San Francisco: Far West Laboratory.

Moody, J. D., & Gifford, V. D. (1990, November). *The effect of grouping by formal reasoning ability levels, group size, and gender on achievement in laboratory chemistry*. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans.

Mugny, B., & Doise, W. (1978). Sociocognitive conflict and structuration of individual and collective performances. *European Journal of Social Psychology*, **8**, 181–192.

Murray, F. B. (1982). Teaching through social conflict. *Contemporary Educational Psychology*, **7**, 257–271.

Newbern, D., Dansereau, D. F., Patterson, M. E., & Wallace, D. S. (1994, April). *Toward a science of cooperation*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Newmann, F. M., & Thompson, J. (1987). *Effects of cooperative learning on achievement in secondary schools: A summary of research*. Madison, WI: University of Wisconsin, National Center on Effective Secondary Schools.

O'Donnell, A. M. (in press). The effects of explicit incentives on scripted and unscripted cooperation. *Journal of Educational Psychology*.

O'Donnell, A. M., & Dansereau, D. F. (1992). Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 120–144). New York: Cambridge Univ. Press.

Okebukola, P. A. (1985). The relative effectiveness of cooperative and competitive interaction techniques in strengthening students' performance in science classes. *Science Education*, **69**, 501–509.

Okebulola, P. A. (1986). The influence of preferred learning styles on cooperative learning in science. *Science Education*, **70**, 509–517.

Palincsar, A. S. (1987, April). *Reciprocal teaching: Field evaluations in remedial and content area reading*. Paper presented at the annual convention of the American Educational Research Association, Washington, DC.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension monitoring activities. *Cognition and Instruction*, **2**, 117–175.

Palincsar, A. S., Brown, A. L., & Martin, S. M. (1987). Peer interaction in reading comprehension instruction. *Educational Psychologist*, **22**, 231–253.

Perret-Clermont, A-N. (1980). *Social interaction and cognitive development in children*. London: Academic Press.

Piaget, J. (1926). *The language and thought of the child*. New York: Harcourt Brace.

Puma, M. J., Jones, C. C., Rock, D., & Fernandez, R. (1993). *Prospects: The congressionally mandated study of educational growth and opportunity*. Interim Report. Bethesda, MD: Abt Associates.

Rich, Y., Amir, Y., & Slavin, R. E. (1986). *Instructional strategies for improving children's cross-ethnic relations*. Ramat Gan, Israel: Bar Ilan University, Institute for the Advancement of Social Integration in the Schools.

Robinson, G. E. (1990). Synthesis of research on class size. *Educational Leadership*, **47**(7), 80–90.

Rosenshine, B., & Meister, C. (1992). The use of scaffolds for teaching higher level cognitive strategies. *Educational Leadership*, **49**(7), 26–33.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, **64**, 4788–530.

Sharan, S., & Hertz-Lazarowitz, R. (1980). A group-investigation method of cooperative learning in the classroom. In S. Sharan, P. Hare, C. Webb, & R. Hertz-Lazarowitz (Eds.), *Cooperation in education*. Provo, UT: Brigham Young Univ. Press.

Sharan, S., Kussell, P., Hertz-Lazarowitz, R., Bejarano, Y., Raviv, S., & Sharan, Y. (1984). *Cooperative learning in the classroom: Research in desegregated schools*. Hillsdale, NJ: Erlbaum.

Sharan, S., & Shachar, C. (1988). *Language and learning in the cooperative classroom*. New York: Springer-Verlag.

Sharan, Y., & Sharan, S. (1992). *Expanding cooperative learning through group investigation*. New York: Teachers College Press.

Slavin, R. E. (1977). Classroom reward structure: An analytic and practical review. *Review of Educational Research*, **47**, 633–650.

Slavin, R. E. (1983a). *Cooperative learning*. New York: Longman.

Slavin, R. E. (1983b). When does cooperative learning increase student achievement? *Psychological Bulletin*, **94**, 429–445.

Slavin, R. E. (1985). Team-Assisted Individualization: Combining cooperative learning and individualized instruction in mathematics. In R. E. Slavin, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb, & R. Schmuck (Eds.), *Learning to cooperative, cooperating to learn* (pp. 177–209). New York: Plenum.

Slavin, R. E. (1987). Cooperative learning: Where behavioral and humanistic approaches to classroom motivation meet. *Elementary School Journal*, **88**, 9–37.

Slavin, R. E. (1989). Cooperative learning and achievement: Six theoretical perspectives. In C. Ames and M. L. Maehr (Eds.), *Advances in motivation and achievement*. Greenwich, CT: JAI Press.

Slavin, R. E. (1991). Are cooperative learning and untracking harmful to the gifted? *Educational Leadership*, **48**(6), 68–71.

Slavin, R. E. (1992). When and why does cooperative learning increase achievement? Theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 145–173). New York: Cambridge Univ. Press.

Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Boston: Allyn & Bacon.

Slavin, R. E. (1994). *Using student team learning* (2nd ed.). Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Slavin, R. E. (in press). Cooperative learning: Theory, research, and implications for active learning. In D. Stern (Ed.), *Active learning*. Paris: Organization for Economic Co-operation and Development.

Slavin, R. E., Madden, N. A., Dolan, L. J., & Wasik, B. A. (1994). Roots and Wings: Inspiring academic excellence. *Educational Leadership*, **52**(3), 10–13.

Solomon, D., Watson, M., Schaps, E., Battistich, V., & Solomon, J. (1990). Cooperative learning as part of a comprehensive classroom program designed to promote prosocial development. In S. Sharan (Ed.) *Cooperative learning: Theory and research*. New York: Praeger.

Stern, D. (Ed.). (in press). *Active learning*. Paris: Organization for Economic Co-operation and Development.

Stevens, R. J., & Slavin, R. E. (1995a). Effects of a cooperative learning approach in reading and writing on academically handicapped and nonhandicapped students. *The Elementary School Journal*, **95**(3), 241–262.

Stevens, R. J., & Slavin, R. E. (1995b). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal*, **32**, 321–351.

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly*, **22**, 433–454.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. *Journal of Educational Psychology*, **83**, 8–16.

Talmage, H., Pascarella, E. T., & Ford, S. (1984). The influence of cooperative learning strategies on teacher practices, student perceptions of the learning environment, and academic achievement. *American Educational Research Journal*, **21**, 163–179.

Van Oudenhoven, J. P., Van Berkum, G., & Swen-Koopmans, T. (1987). Effect of cooperation and shared feedback on spelling achievement. *Journal of Educational Psychology*, **79**, 92–94.

Van Oudenhoven, J. P., Wiersma, B., & Van Yperen, N. (1987). Effects of cooperation and feedback by fellow pupils on spelling achievement. *European Journal of Psychology of Education*, **2**, 83–91.

Vroom, V. H. (1969). Industrial social psychology. In G. Lindzey and E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., Vol. 5). Reading, Ma: Addison-Wesley.

Vygotsky, L. S. (1978). *Mind in society* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.), Cambridge, MA: Harvard Univ. Press.

Wadsworth, B. J. (1984). *Piaget's theory of cognitive and affective development* (3rd ed.). New York: Longman.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, **13**, 21–39.

Webb, N. M. (1992). Testing a theoretical model of student interaction and learning in small groups. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 102–119). New York: Cambridge Univ. Press.

Wheeler, R., & Ryan, F. L. (1973) Effects of cooperative and competitive classroom environments on the attitudes and achievement of elementary school students

engaged in social studies inquiry activities. *Journal of Educational Psychology*, **65**, 402–407.

Wittrock, M. C. (1986). Students' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) New York: Macmillan.

Yager, S., Johnson, R. T., Johnson, D. W., & Snider, B. (1986). The impact of group processing on achievement in cooperative learning. *Journal of Social Psychology*, **126**, 389–397.

86

COOPERATIVE LEARNING IN CLASSROOMS

Processes and outcomes

N. Bennett

This article focuses on the description and analysis of our research on cooperative grouping over the last 8 years. This research effort has moved through three interconnected phases—from description of classroom practice, to experimentation, to implementation.

Descriptions of typical classroom practice have established the paucity of cooperation. Groups tend to be no more than collections of children sitting together but engaged on individual work. In such groups the level of cooperation, frequency of explanations and knowledge exchange is low.

Thus, in order to gain an understanding of the nature and process of cooperation in groups on normal curriculum tasks required the setting up of within-classroom experiments. These revealed, among others, that group composition is important to learning outcomes, and that pupil involvement substantially improves in cooperative group endeavours.

Missing, however, were data on the successful implementation of cooperative groups. Yet successful implementation has great contemporary relevance because of the demand for assessed skills in collaboration in the national curriculum. Our current work therefore addresses the impact of changing grouping practices on group processes, classroom management, the teacher's role and children's learning.

Preliminary findings from this study are presented together with a wider analysis of the state of the research field and possible ways forward.

Introduction

It might be supposed that those skills and abilities which have made man a successful species would both be valued and developed in schools. However,

Source: *Journal of Child Psychology and Psychiatry*, 1991, 32(4), 581–594.

Schmuck (1985), commenting on school practices in the United States, argues otherwise. In responding to the question 'Why have we humans been so successful as a species?', he argued as follows.

> "We are not strong like tigers, big like elephants, protectively coloured like lizards, or swift like gazelles. We are intelligent, but an intelligent human alone in the forest would not survive for long. What has really made us such successful animals is our ability to apply our intelligence to co-operating with others to accomplish group goals. From the primitive hunting group to the co-operative boardroom, it is those of us who can solve problems while working with others who succeed. In fact, in modern society, co-operation in face-to-face groups is increasingly important . . . It is difficult to think of very many adult activities in which the ability to co-operate with others is not important.
>
> Because schools socialize children to assume adult roles, and because co-operation is so much a part of adult life, one might expect that co-operative activity would be emphasized. However, this is far from true. Among the prominent institutions of our society, the schools are least characterized by co-operative activity."

But is this same picture true of British primary schools whose teaching practices are underpinned, in theory at least, by an ideology which stresses democracy and participation? And if it is, then what actually passes for cooperation, and how might it be improved? These are some of the questions to which we have been seeking answers in a series of studies extending over the last 7–8 years. Here I aim to reflect on these studies, chart our progress and consider possible ways forward.

My research interests over the last two decades have centred around gaining better theoretical and practical understandings of teaching–learning processes in natural classroom settings, particularly at primary school level. This is an area with a long, but relatively inglorious history, so that as late as the mid-1970s one reviewer was writing, "The research on teaching in natural settings to date has tended to be chaotic, unorganised and self-serving . . . There seems to be no simple route through the chaos which has developed" (Rosenshine & Furst, 1973). One suggested way of imposing order on this chaos was for researchers to utilize a research model called the descriptive–correlational–experimental loop. It has three elements:

(1) development of procedures for description in a quantitative manner;
(2) correlational studies in which descriptive variables are related to pupil variables;
(3) experimental studies in which the significant variables obtained in the correlational studies are tested in a more controlled situation.

Our work on cooperative grouping has essentially followed this model. We began by carefully describing typical classroom practice, then moved to a quasi-experimental phase in which we considered the relationship between group processes and outcomes, before moving into an intervention or implementation stage in which we further investigated group processes alongside a consideration of implementation issues such as task design, classroom management and assessment.

Grouping practices in classrooms

My initial interest in groups came from studying teaching and learning in classrooms from a contructivist perspective, utilizing insights derived from cognitive psychology and from models of mediating social processes (Bennett & Desforges, 1988). This perspective assumes that the tasks on which pupils engage structure to a large extent what information is selected from the environment and how it is processed. The classroom task is the most proximal point of contact between teacher and pupil with respect to the pupil's organization and knowledge of curricular areas. Thus to understand classroom learning requires an understanding of children's progressive performances on the tasks they are assigned (or choose). Further, and most imporant in this context, since classroom learning takes place within a complex social environment, it is necessary to understand the impact of social processes on children's task performances.

The immediate social context of learning in primary classrooms in Britain is a small group typically containing between four and six children (HMI, 1978; Bennett, Andreae, Hegarty & Wade, 1980). The initial research question we posed was, 'does working in such a group aid or hinder the performance of tasks?' At that time, in the early 1980s, there had been little research on grouping practices. Professional commentary from Her Majesty's Inspectorate had suggested that there was little cooperative grouping to be seen, and the limited observational studies which had been carried out were of doubtful validity. These presented stories of low pupil involvement in groups, and marked sex differences whereby boys tended not to talk to girls and vice versa. Finally there was an uncompromising finding that children worked *in* groups, not *as* groups. Less than 10% of all observations were of groups working cooperatively (Galton, Simon & Croll, 1980).

This picture was far removed from official prescription. The Plowden Report (1967), for example, encouraged grouping, perceiving it to have clear benefits for children's social and intellectual development and as providing the best compromise in achieving individualization of learning and teaching within the time available. "Sharing out the teacher's time is a major problem. Only seven or eight minutes a day would be available for each child if all teaching were individual. Teachers therefore have to economize

by teaching together a small group of children who are roughly at the same stage" (para. 754/5).

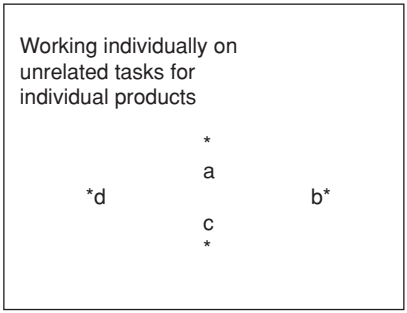
The approach adopted in our first study was a combination of observation and audio-recording of children's talk in groups, in classrooms of different socio-economic intake, taught by teachers regarded as better than average by their peers. Accordingly we recorded and transcribed all the talk of children aged between 6 and 7, in language and maths tasks. This talk was categorized using a scheme grounded in the data, i.e. derived from successive readings of the transcripts. The quantitative and qualitative analyses presented, in some respects, a more positive picture than previous studies (Bennett, Desforges, Cockburn & Wilkinson, 1984).

Three-quarters of all talk was task-related. However, when this task-related talk was analysed further it became clear that little of it was task-enhancing, i.e. helped in the performance or understanding of the task. Talk tended to be about procedure, discussions of how much work each child had completed, talk of each other's social or intellectual competence and so on. The category of talk most likely to be task-enhancing was one we called 'instructional input', where children shared knowledge and provided, or received, explanations. Only 16% of all talk appeared in this category, although there was a great variation across the groups studied. In view of its prospective importance all talk in this category was analysed qualitatively. The results were sobering. The great majority of requests and responses were of low order: a specific question, for example. "How many twos in 54?", followed by a specific response, and not all responses were correct. Explanations were rare, as indeed were sex differences at this age.

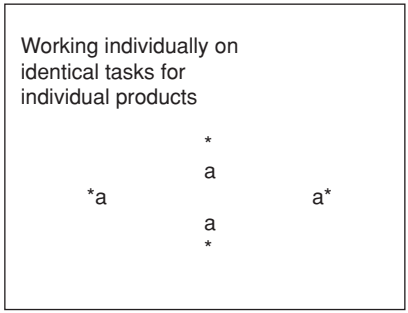
Analyses of profiles of talk in groups of differing ability showed a wide variation. In groups containing high and average ability pupils it was found, for example, that instructional input was substantially greater than that found in groups composed of low and average ability pupils. Some light was also thrown on teacher management issues. Teacher talk to groups in this sample was minimal. The great majority of teacher–pupil interactions took place at the teacher's desk as work was marked. This arose as a consequence of teacher's decisions that work must be marked with the pupil present. The implementation of this decision had wider organizational consequences, however, leading to constant queuing, lack of group teaching and no group supervision, the effect of which is likely to be a significant diminution in task-related behaviour. American studies have found, for example, that pupil involvement in groups is high whilst substantial instructional interaction is taking place, but drops to around 50% when the teacher is not supervising the group or not interacting with pupils in the group context (Fisher *et al.*, 1978). The typical management style observed on this study ate up teacher time, making it impossible for them to adequately monitor group work or indeed give adequate time to any individual child.

The grouping approach of the teachers studied was typical of those reported by other researchers. Children worked in a group but not as a group. They sat in physical grouping but worked on individual tasks. Cooperation was neither overtly encouraged nor discouraged, leading to confusion among a number of children who were clearly unsure whether or not it was allowed.

These non-cooperative grouping models are shown schematically below.



Scheme (i) * = children; a, b, etc. = tasks



Scheme (ii) * = children; a = tasks

Scheme (i) is typical in maths where children are working on a structured scheme but are at different stages. This even occurs when children sit in homogeneous ability groups. So, each of the children work on different, i.e. individual, tasks, each aiming for an individual end product.

Scheme (ii) is more common in writing where the teacher's request is often for a class task, e.g., the writing of a story. In this case all the children are engaged on identical work.

What seems to have happened in practice is that teachers have taken on board the Plowden Report's views on having children work in groups, but have preferred to retain individualization rather than cooperation in that context. As one educationalist wrote:

> "Grouping thus emerges as an organisational device rather than as a means of promoting more effective learning, or perhaps exists for no reason other than that fashion and ideology dictate it" (Alexander, 1984).

Cooperative grouping

These findings, together with those of earlier studies, showed that all was not well in the social context of classroom learning. Yet paradoxically, increasing evidence was emanating from the United States at that time of the social and academic benefits of cooperative groupwork (cf. Slavin, 1983). One meta-analysis concluded that the strong positive effects of cooperative

grouping across subjects and age levels "stand as strong evidence for the superiority of cooperation in promoting achievement and productivity . . . educators may wish to considerably increase the use of cooperative learning procedures to promote student achievement" (Johnson, Manruyama, Johnson, Nelson & Shaw, 1981).

However, most of these American studies were limited to considerations of input and output, showing scant regard for *how* groups were effective. Studies of how group processes relate to outcomes have been few in number and limited in scope. They were, nevertheless, important in informing our later studies (cf. Bennett, 1987).

The variable most studied in American research has been helping behaviour, not dissimilar to our 'instructional input' category. It has been found that giving and receiving help both relate to achievement, although to be most effective these have to be in the form of explanations rather than in the form of telling. High attaining children are the main source of giving help (Webb, 1989). A major limitation of this American work, however, was that the influence of group composition on group processes and outcomes had not been established.

The results of our first study, together with the American evidence, argued for an investigation of the quality of group processes and outcomes in groups of differing composition engaged on cooperative tasks. In this, our second study, observation and audio-recording of talk in homogeneous, heterogeneous and mixed groups were supplemented by individual interviews with each child to ascertain their recall and understanding of the cooperative decisions made and their explanations and justifications for them. A category system for classifying group talk was again derived from the data, the resultant categories being grouped into two major categories of 'instructional talk' and 'procedural' or 'management talk' (Bennett & Cass, 1989).

Analyses revealed that within-group differences were of most interest. In the homogeneous groups of all high, all average and all low ability children, the all high group consistently and significantly out-performed the average and low ability groups. This group talked more, and more of that talk was instructional, they provided three-quarters of all the explanations, they had the best recall of decisions taken and gave the most valid reasons for those decisions.

The low and average attaining groups on the other hand were characterized by a very low level of instructional talk and virtually no explanations. The low group for example gave just five of the 155 explanations sampled in this study. Also both groups gave a proportionately larger number of incorrect reasons for decisions.

The within-group differences in the mixed groups were also of interest and potential value. Within the 'mixed' groups of three children were two variants—two low and one high attainer, and two high and one low attainer. On every criterion it was the two low and one high group which was

superior. What appeared to happen in the other combination is that the two high attainers talked together whilst the low attainer was ignored, or opted out, and as a consequence misunderstood the basis on which decisions were being made. In the two low and one high combination, on the other hand, the high attainer took on the role of peer tutor and support.

When considering the effect of attainment level, irrespective of group membership, we found that high attaining children performed well irrespective of the group they happened to be a member of. This is an important finding since many teachers appear to fear that grouping such children with lower attainers adversely affects their progress. Sex differences on the other hand were not large. Girls tended to talk slightly more, they gave fewer explanations but more of these were correct, and similarly gave less reasons for the decisions made, but more were correct.

Other findings which support earlier work, and the value of cooperative groupwork, are that task involvement was extremely high, a very low level of off-task behaviour was observed, and the ratio of instructional to procedural talk was also high. Finally, there was evidence of a link between general verbal participation in the group and achievement.

Although no great claims for the external validity of this study were made, it raised several important issues. Firstly, group composition is important, both in terms of the quality of the processes within the group, and outcomes. Particularly worrying is the very poor performances of average and low ability groups. These findings have clear implications for teachers' classroom management strategies since the majority of teachers appear to group by ability. In this structure high attaining groups will perform well but the low and average groups will need substantial teacher support to be successful. Secondly, that high attainers perform well irrespective of their group membership is important because many teachers worry that grouping such children with low attainers adversely affects their progress. Thirdly, some links were established between group processes and outcomes in cooperative classroom groups, and finally, there was evidence of very high pupil involvement in cooperative settings.

These findings generally support the claims made for cooperative groupwork elsewhere (Webb, 1989). However, by controlling the task demands and the management of the task, this study, like all others, was unable to provide any information on the critical issues of implementing cooperative grouping in classrooms, and the teacher's role in that process.

Implementing cooperative groupwork

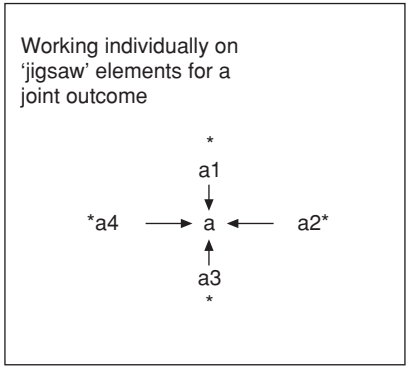
The need for research on implementation has attained great contemporary relevance in Britain as a consequence of the introduction of a national curriculum which is demanding assessed skills in cooperative endeavours. For example, the English curriculum 'speaking and listening' attainment targets

demand that by the age of 7 children should be able to present real or imaginary events in a connected narrative to a small group of peers, speaking freely and audibly; and by the age of 11 the average child will be able to offer reasoned explanations of how a task has been done or a problem solved and to take part effectively in small group discussion. The particular pedagogical problem that our research has highlighted is that the national curriculum is making demands for cooperative groupwork from a teacher work force which has little experience, or expertise, in this kind of teaching.

An allied problem for the primary school teacher is the need, under the national curriculum, for the continuous assessment and recording of children's performances. This demand marks a major change in practice from informal, unrecorded assessment to the opposite. It will undeniably require more time, particularly if teachers heed the exhortation from ourselves and others that assessment should be diagnostic. Carefully ascertaining children's understandings and misconceptions of content and concepts is far more time-consuming than marking a page of sums. A constant cry from teachers is therefore, "where is the time to be found?" We believe that this time can in fact be found by altering the authority structure between teacher and group.

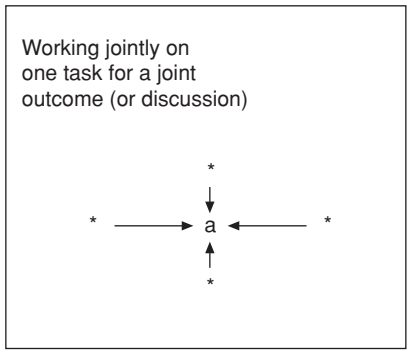
The major aim of our present study is thus to investigate the impact of changing classroom grouping practices from those requiring individual, to those requiring cooperative, outcomes on such aspects as group interaction, classroom management, teacher activity and the creation of teacher time. A particular focus is on the choice and design of group tasks in order that we might ascertain relationships between task demand, types of group talk and pupil outcomes.

We have been working with a group of 15 teachers who, between them, teach the entire primary age range from 4 to 12 years. All were utilizing typical, i.e. non-cooperative, grouping practices, but agreed to shift to one of two cooperative group models. These models are shown below.



Model 1 * = children; a1, a2, etc. = tasks

In the kind of task shown in Model 1 (modified from Aronson, 1978), there are as many elements to the task as there are group members. Each child works on one element and the task is divided in such a way that the group outcome cannot be achieved until every group member has successfully completed his piece of work. At this point the 'jigsaw' can be fitted together. Cooperation is thus built into the task structure, as indeed is individual accountability. It is difficult in this type of group task for a child to sit back and let others do all the work, especially since group members are likely to ensure that everyone pulls his/her weight. Examples of such tasks would be the production of a group story or newspaper, or the making of a 'set' of objects in a practical maths activity.



Model 2 * = children; a = tasks

For the type of task shown in Model 2 (derived from Johnson & Johnson, 1975), children will need to work cooperatively since only one product will be required of the group. Activities will therefore have to be coordinated and it is possible that a group leader will emerge in order to create the necessary organization. Each individual's work will have an impact on the group product but will be worthless until it becomes part of that product. Examples can be seen in problem solving in technology, construction activities or in discussion tasks. Although collaborative endeavour is necessary for the group to succeed, it is less easy to ascertain exactly what each group member has contributed and individual accountability is therefore lower.

The teachers cooperatively designed the tasks, in the areas of maths, language, science and technology. They implemented the cooperative group models in their classrooms over a 6-week period during which they also created a new authority structure via a classroom rule that all pupil demands and requests must first be dealt with by the group. Only when this group responsibility procedure had been exhausted could pupils approach the teacher.

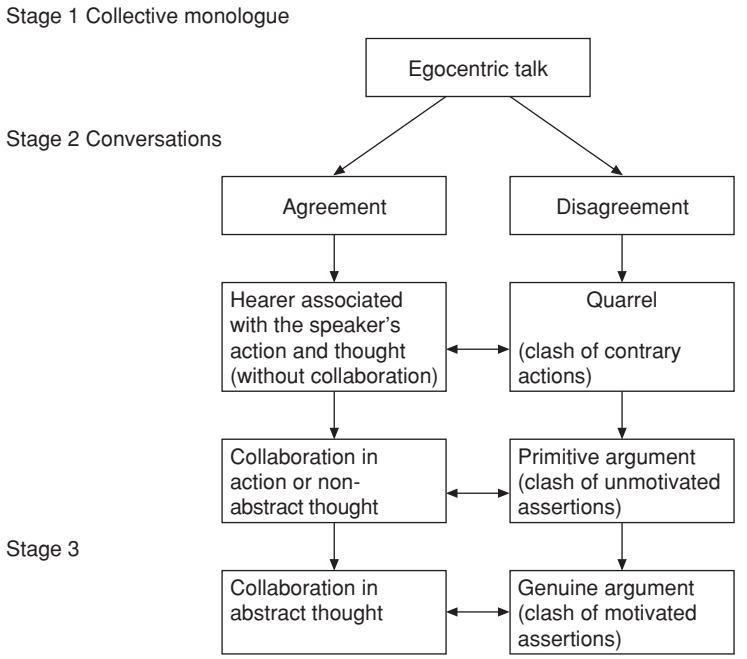


Figure 1 The development of children's conversations

Tasks were recorded and transcribed and observations made of teacher activity, classroom management strategies and pupil demands. Teachers also provided reflective accounts of the implementation. They also provided their professional judgments of the quality of children's work.

As indicated earlier one of the major foci of the study is the relationship between task demand and type of talk, since this is a central issue in understanding the links between teaching and learning. In order to fulfil this foci, an analytic system was required which was capable of generating conceptual links between types of group talk and types of task demand. The conceptualization we found most useful in this regard was Piaget's (1959) model of the development of children's conversations. The stages he proposes for children aged 4–7 are shown in Fig. 1.

Briefly, Piaget argues that after children have progressed through the egocentric stage various types of interaction or conversation are apparent, based on whether the interaction is in agreement or disagreement, and on differing levels of collaboration in action and abstract thought. It is only within conversation at stage 3 that "there is any real interchange of thought".

Our concern is not with developmental aspects of this model; it is simply being used as a heuristic to postulate the seven modes of group interaction

Table 1 Conversational modes

	Stage	*Conversational mode*	*Mode*
	1	Collective monologue	A
Action	2	Association with	B
		Sharing in	C
		Collaboration in	D
		Quarrelling	E
		Primitive argument	F
Abstract thought	3	Collaboration	G
		Genuine argument	H

Table 2 Conversational modes in language and maths (%)

Mode	*Language*	*Maths*
A	6	11
B	10	52
C	43	26
D	14	3
E	2	7
F	2	
G	21	
H	3	

shown in Table 1 (for definitions and examples see Bennett & Dunne, 1989).

These conversational modes are consistent with Piaget's model as are the definitions of the terms Action and Abstract.

Action is talk related to the activity of the moment.

Abstract is talk no longer connected with the activity of the moment, but concerned with finding an explanation, reconstructing a story or a memory, discussing the order of events or the truths of a tale.

Analyses of talk by conversational modes has allowed us to consider more coherently than before the relation between group talk and task demand. So far we have only analysed the language and maths tasks and a comparison of these by conversational mode shows up an interesting trend (Table 2).

Marked differences in the proportions of conversational modes are apparent in the two areas. Talk in maths is heavily concentrated in mode B—'associating with', i.e. talking about, or commenting on, their own activity, and thus is conversational not collaborative. There is no evidence of any abstract talk at all in maths.

In contrast, one-quarter of the talk in language work is abstract, and the talk relating to action is more sophisticated. Much more is in mode C—'sharing in', talk centring around a shared activity often involving demonstration or a response/request for help.

Thus, maths tasks can be characterized as demanding action talk, whereas language tasks demand both action and abstract talk.

The reason for this dichotomy seems to be fairly clear: in maths and science/technology the tasks set for the children, across the whole age range, could be defined as action tasks. Pupils were asked, for example, to make cubes, to make triangular prisms, to make carts that will roll down a slope, to make models for a fairground. These tasks are practical and involve manipulation of materials. The children *have* to be involved in action in order to complete their tasks.

On the other hand, language tasks generally demand a different kind of activity. The children are asked to talk, to discuss, to make decisions unrelated to action. They are given problems to be solved verbally, tasks which ask them to look for meaning, to provide ideas, to compare and contrast. Alongside this, there is a demand to write or draw, sometimes as a group response, sometimes as individuals, in order to show what has been gained from the group discussion. Such tasks are characterized by a predominant demand for abstract thought, the action demand being secondary. So, for example, the whole group provide ideas for a single story, the emphasis is on creative thought; or the teacher provides a written or drawn stimulus for discussion, sometimes with questions to guide thinking. There is then a demand that this be completed by writing or by drawing.

When tasks combine action with abstract demands, the talk related to action dominates. It seems as if, given the opportunity to talk about action, the children will take it. Or it may be that, since an end-product is always demanded by the teacher, the action required for this is given the group's greatest attention. Thus, in language lessons where there has been a real attempt verbally to solve a problem as a group, talk relating to the writing up, or to the drawing of ideas and meanings, is always proportionately higher.

There is also a relationship between group type and conversation mode. It seems that when children are asked to submit a group response, rather than several individual ones, the amount of talk relating to action is significantly diminished. Ironically, this has occasionally led to teachers suggesting that such a lesson has been less worthwhile, since there has been less talk when compared with other sessions, and sometimes more off-task chatter; in fact, it may be instead that abstract talk is harder, comes less easily to the children, particularly at this age, needs more pause for thought, and that their participation is less obvious.

The proportion of pupil involvement or on-task talk in these cooperative groups was again significantly higher than for typical classroom groupings.

Table 3 Task-related talk during individual and cooperative group work

	Cooperative group work (Bennett & Dunne, 1989)	*Individualized work (Bennett et al., 1984)*
Maths	89%	63%
Language	86%	70%
Total	88%	66%

Table 3 compares the proportion of task-related talk in our earlier study of typical, non-cooperative groupings with those of the current study, where it becomes clear that the differences in favour of cooperative groupings are substantial.

And what of the teachers' responses? Did they find difficulties in managing their new systems? Did the change in authority structure actually create time? Were they satisfied with the quality of their pupils' work? The short answer to these questions is yes.

For all these teachers, cooperative grouping was entirely new, but without exception they found it easier to implement than they had imagined. Typical comments were "I was pleasantly surprised at how easy the sessions were", and, "The children performed in a more business-like way than I'd expected". Most of the teachers also commented on how much the children enjoyed such group work and how enthusiastic they were to continue to work in this way. One initially sceptical teacher said she was "agreeably surprised to find that the children were in fact able to use each other and help each other more than I realized".

The freeing of teacher time is important, and here too the results have been encouraging. We know, from our own observations, that time was created in each classroom, and the teacher comments bear this out. "Much more time is available to teach rather than to deal with many matters which can be peer assisted." Another reinforced this, stating that, "It is a management method that really frees the teacher, and would enable her to carry out the profiling, observation and testing jobs". So much time was freed in some classrooms that the teachers began to feel guilty at not being rushed off their feet—"I found it very satisfying teaching in this way because the children were so involved in their work. It gave me a lot of free time . . . At times this made me feel that I was not doing my job". Not only did the frequency of demand diminish but the type of demand changed. The questions which pupils eventually brought to teachers were of a higher level.

The extent and quality of children's involvement was commented on by all teachers who saw, for example, "a dramatic increase in the amount of discussion, suggestion, testing, inferring and drawing informed conclusions"

and "rich mathematical language". They reported "work that was more thorough and presented well" and children who were "thinking and reflecting their views and not a teacher's". One stated of a poetry activity: "I was delighted that this whole assignment had developed the conditions for such high quality learning to take place".

Low attainers particularly seemed to benefit and in some classes they "now often sit with higher ability children who took them under their wing". On the other hand we have evidence to show how much high attainers gain from group work.

Overall, both the teachers with whom we have worked in the past, and those with whom we continue to work, are positive about the benefits of the approach, seeing it as beneficial to themselves in terms of releasing teacher time, and beneficial to children in terms of greater independence, greater cooperation and better quality work from both low and high attainers.

Conclusion

Our studies have shown both the aridity of typical classroom grouping practices and the promise of cooperative approaches. Nevertheless, I do not want to be perceived as promulgating a panacea. I do not prescribe it as *the* method. What I do advocate is a better balance of teaching approach than at present between individual, group and whole class teaching, with grouping perhaps taking a pre-eminent role in problem solving and application tasks.

Acquiring a better balance of teaching approaches will require support for teachers. Many a good idea has foundered on the rocks of implementation, and rocks there are. Changes are necessary in the ways in which teachers plan and develop tasks, and in the ways in which they organize and manage their classrooms. We do not underestimate these difficulties and have, towards this end, recently written a workbook for teachers who wish to implement cooperative grouping (Dunne & Bennett, 1990).

There is, inevitably, more research to be done. We have gained a better understanding of the inter-relationships of group processes and outcomes, and of the mediating roles of talk, tasks and group composition. The distinction between action and abstract talk has generated another set of questions. Is conversational mode mediated by variables other than task demand, such as group type, composition or age of children? Can we design tasks in maths, science and technology that generate abstract talk? Can children be trained to function more effectively in groups, at both a social and intellectual level? What should the nature of teacher intervention be in cooperative group work?

Our initial interest in groups developed as a direct consequence of our work on the role of task structures and classroom management on the quality of children's learning experiences. We retain these interests, but our studies have drawn our attention increasingly to the importance of talk in

children's learning and in particular to the theoretical ideas of Vygotsky about the links between learning and social processes. He believed, for example, that "learning awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in his environment and in co-operation with his peers" (Vygotsky, 1978, p. 90). It is apt, therefore, that I conclude with another quote from Vygotsky: "A word devoid of thought is a dead thing, and a thought unembodied in words remains a shadow". For him, this sentence symbolized the relationship between thought and speech: for me, it symbolizes the overarching aim of our work on groups—to bring thought and talk out of the shadows of British primary classrooms.

References

Alexander, R. (1984). *Primary teaching*. London: Holt Rinehart Winston.

Aronson, E. (1978). *The jigsaw classroom*. Beverly Hills: Sage.

Bennett, N. (1987). Cooperative learning: children do it in groups—or do they?! *Educational and Child Psychology*, **4**, 7–18.

Bennett, N., Andreae, J., Hegarty, P. & Wade, B. (1980). *Open plan schools*. Windsor: NFER.

Bennett, N. & Cass, A. (1989). The effects of group composition on group interactive processes and pupil understanding. *British Educational Research Journal*, **15**, 19–32.

Bennett, N. & Desforges, C. (1988). Matching classroom tasks to students' attainments. *Elementary School Journal*, **88**, 221–234.

Bennett, N., Desforges, C., Cockburn, A. & Wilkinson, B. (1984). *The quality of pupil learning experiences*. London: Erlbaum.

Bennett, N. & Dunne, E. (1989). Implementing co-operative groundwork in classrooms. Paper presented at EARLI Conference, Madrid, September 1989.

Dunne, E. & Bennett, N. (1990). *Talking and learning in groups*. London: Macmillan.

Fisher, C. W., Filby, N. N., Marliave, R., Cohen, L. S., Dishaw, M. M., Moore, J. E. & Berliner, D. (1978). *Teaching behaviours, academic learning time and student achievement* (Final Report), *Beginning teacher evaluation study*. San Francisco: Far West Lab.

Galton, M., Simon, B. & Croll, P. (1980). *Inside the primary classroom*. London: Routledge & Kegan Paul. HMI (1978). *Primary education in England*. London: HMSO.

Johnson, D. W. & Johnson, R. (1975). *Learning together and alone: cooperation, competition and individualisation*. Englewood Cliffs: Prentice-Hall.

Johnson, D. W., Manruyama, G., Johnson, R., Nelson, D. & Shaw, L. (1981). Effects of co-operative, competitive and individualist goal structures on achievement: a meta-analysis. *Psychological Bulletin*, **89**, 47–62.

Piaget, J. (1959). *The language and thought of the child*. London: Routledge & Kegan Paul.

Plowden Report (1967). *Children and their primary schools*. London: HMSO.

Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally.

Schmuck, R. A. (1985). In R. E. Slavin (Ed.), *Learning to cooperate, cooperating to learn*. New York: Plenum.

Slavin, R. E. (1983). *Cooperative learning*. New York: Longman.

Vygotsky, L. S. (1978). *Thought and language*. Cambridge: MIT Press.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, **13**, 21–39.

87

COOPERATIVE LEARNING AND PEER TUTORING

An overview

K. Topping

We live in an increasingly individualistic, competitive society. Government requirements set children against children and schools against schools. Given the prevailing atmosphere of Victorian Social Darwinism, the assumption that upward mobility can only be at the expense of others is readily made. In such a climate, can cooperative learning possibly have any viable role to play?

Definitions of cooperative learning

As with many other concepts in the wide field of education, cooperative learning may seem a nebulous and elusive notion. Leaving aside the problematic issue of a consensual definition of "learning", it is instructive to study the roots: CO- means together, in company, jointly, in common, equally, mutually, reciprocally, while -OPERATE means to work, act, influence, effect, accomplish, cause or carry out. From these options 49 definitions of "cooperative" may be generated – enough for a start.

Cooperative learning is however more than "working together". It implies synergy, a combined action of differentiated specialists, as in a symbiosis of bodily organs or mental faculties. Cooperative learning methods are about "structuring positive interdependence" (Davidson, 1990).

Existing practice

Of course, teachers have been doing it for years. However, they have not been doing it as much as they might think, nor as effectively. Doyle (1986) noted that certainly into the early '80s in the USA small groups in which students worked together on assignments were used infrequently in most

Source: *The Psychologist*, 1992, 5, 151–157.

classrooms. The promotion of structured approaches to cooperative learning has since gathered pace in the USA, but remains something of a minority interest.

In the UK, the practice of grouping was advocated by the Plowden Report (1967), which embodied the assumption that such methodology represented a cost-effective use of teacher time. Subsequently, a survey by Her Majesty's Inspectorate (1978) showed that most primary school teachers grouped children for some aspects of the curriculum, and in basic skills learning homogenous ability grouping was the most common format. Mixed ability or friendship grouping was more widely used for Science, Art and Crafts. The assumed benefits of small groups were restated in subsequent HMI reports through until the middle eighties when the ideology began to change in line with current political imperatives.

In the late 1980s the introduction of the GCSE with its emphasis on continuous assessment facilitated more group work, which the previous examination system had militated against. However, as Cowie and Rudduck (1988) noted, these opportunities were not necessarily taken up, much individualised work remaining in examination syllabuses and group work being more prominent in new rather than traditional subject areas.

In any event, as Bennett (1987) observed, research on the utility and functioning of groups in classroom settings provided little support for the benefits assumed by Plowden or HMI. While children sat in groups, for the majority of the time they worked as individuals on their own task. Only one-sixth of the time was spent interacting with other pupils, and most of this was not related to the task. Pupils might be placed in groups to work, but that did not mean that they worked as groups, unless "positive interdependence" was carefully and properly structured.

When talk between pupils in working groups was analysed it emerged that little of this enhanced the task in hand (Bennett *et al*., 1984). Furthermore, teachers often failed to behave in ways which facilitated or encouraged truly cooperative interaction. There are also indications in the literature that while affective outcomes may be positively associated with more open forms of classroom functioning, the findings for cognitive gains are the reverse, and indeed younger children show preference for a more traditional learning environment, while problem students are better adjusted in classes they perceive as high in order and organisation (MacAulay, 1990). Workers such as Wheldall and his associates (1981) have replicated findings that seating students in rows has a positive effect on classroom behaviour and attention levels.

Evidently cooperative group work is unlikely to be actually happening effectively in UK classrooms nearly as much as professional educators might assume, even though, as Rudduck and Cowie (1988) note, it has "many guises". These authors offer a typology of formats: discussion, problem-solving tasks, production tasks, simulations and role-play activities, together with three

more specific techniques: Buzz groups, Snowball groups and Cross-over groups. Davidson (1990) offers an alternative framework, proposing that structuring positive interdependence essentially involves the specification of goals, tasks, resources, roles and rewards.

Objectives of cooperative learning

As with any educational intervention, objectives may be specified in the cognitive, social, affective or other domains. However, the educational "Best Buy" in one domain may not prove effective in another. Certainly, the Plowden assumption that group work equalled cost-effective deployment of teacher time merits more detailed scrutiny. On the basis of a meta-analysis of more than 500 studies of cooperative learning, the Johnson brothers claimed that *structured* cooperative group work was superior to other methods in promoting higher student achievement (Johnson *et al*., 1981; Johnson & Johnson, 1989). However, commentators have regarded this conclusion as far too sweeping, noting the existence of interaction effects with variation in tasks and measures and expressing concern that many of the studies involved had been carried out in laboratory settings rather than naturalistic classrooms (Cotton & Cook, 1982; McGlynn, 1982).

The attraction of cooperative learning methods for many practising teachers rests however on assumptions of impact in the social and affective domains – even if children do not need to learn the curriculum material, they need to learn how to cooperate. This view is encouraged by research such as that of Johnson and Johnson (1983), who in a study of relationships between handicapped and non-handicapped students found that cooperative learning experiences, as compared with competitive and individualistic ones, promoted more interpersonal attraction between handicapped and non-handicapped students and promoted higher self-esteem on the part of all students. However, if cooperative learning methods are to compete (sic) in the educational market place, they must be able to demonstrate cost-effectiveness across all domains – the whole must perforce be considerably more than the sum of its parts.

In this context, Greenwood *et al*. (1990) offer a useful review of research comparing teacher- versus peer-mediated instruction. A number of studies cited by these authors have reported significantly greater gains for cooperative learning groups than for parallel groups receiving traditional teacher instruction. A few studies (eg. Russell & Ford, 1983) have found that cooperative learning methods (usually one-to-one peer tutoring) had greater effects than *supplementary* instruction by a resource teacher in a small group setting. This finding echoes that of Tizard *et al*. (1982) in the field of parental involvement in reading, where parent-tutored children out-performed children receiving additional professional tuition, the differences remaining discernible at follow-up four years later.

Also concerning comparative cost-effectiveness, Sherman and Harris (1975) conducted a series of studies suggesting that assigning time for independent study had a small impact on subsequent classroom performance, prescribing homework had a somewhat greater but erratic effect on performance, while peer tutoring had the most substantial and most consistent effect on classroom performance. Levin and Meister (1986) compared the relative effectiveness of cross-age peer tutoring, reduced class size, a longer school day and computer-assisted instruction. These authors noted that computer-assisted learning is often assumed to be highly cost-effective, but concluded from the evidence that while such an approach was more cost-effective than reducing class size or increasing instructional time, it was less cost-effective than peer tutoring in both mathematics and reading.

Greenwood *et al.* (1990) assert that cooperative learning can result in high rates of time on task, monitoring and feedback, features known to be related to successful learning outcomes, but equally observe that the procedure is not without its costs or ethical concerns. While peer-mediated instruction might be higher on engaged time, opportunities to respond and opportunities for error correction than teacher-mediated instruction, and offer greater immediacy of error correction and more opportunities for help and encouragement, there are onerous requirements in peer training and quality control, while content coverage might be variable and adaptations of curriculum materials necessitated.

USA research on cooperative learning

In the United States, the worker whose methods are most widely discussed is probably Robert Slavin. Highly productive in instructional design, research and dissemination, Slavin has written widely, the most useful current compendium being Slavin (1990). Herein a typology of cooperative learning is offered which emphasises six principal characteristics: group goals, individual accountability, equal opportunities for success, team competition, task specialisation and adaptation to individual needs. Slavin proposes that individual accountability may be achieved by having group outcome scores represent the sum or average of individual quiz scores or other assessments. This, and the use of competition between teams as a means of motivating students, would be alien to the inclinations of some UK teachers.

Slavin's highly structured procedures such as STAD (Student Teams – Achievement Divisions), TGT (Teams – Games – Tournaments), JIGSAW (I and II), TAI (Team Assisted Instruction) and CIRC (Co-operative Integrated Reading and Composition) have been well researched. A meta-analysis (Slavin, 1990) yielded effect sizes of 0.27 for STAD, 0.38 for TGT, and 0.21 for TAI/CIRC but of only 0.04 for JIGSAW. In comparison, other more loosely structured forms of cooperative learning yielded effect sizes ranging from 0.00 to 0.12, and further analysis indicated that the higher effect sizes

tended to be associated with approaches which combined group goals and individual accountability. Slavin commented "traditional group work, in which students are encouraged to work together but are given little structure and few incentives to do so, has been repeatedly found to have small or non-existent effects on student learning" (pages 30/31). Slavin (1990) described the effect sizes associated with his own procedures as "moderate but important effects", particularly since they could (he claimed) be achieved in practice at very little cost. He commented that "effect sizes of +0.25 are generally considered educationally significant" and noted that few educational interventions had produced effect sizes much larger, with the exception of one-to-one tutoring, whilst perhaps surprisingly failing to make reference to peer tutoring in this context.

By contrast, the Johnson brothers have placed less emphasis on extrinsic reward and team competition and more on the internal structure of the cooperative learning activity (eg. Johnson & Johnson, 1991; and the aforementioned meta-analyses in Johnson *et al.*, 1981, and Johnson & Johnson, 1989), demonstrating a greater interest than Slavin in social and affective outcomes.

Davidson (1990) provided a summary of studies of achievement outcomes comparing small cooperative group instruction with expository instruction or individualised instruction in mathematics. Of 72 studies, 36 demonstrated a statistically significant difference in favour of the small group method, and it was evident that significant results were more frequently found with Slavinesque structures than with other forms of organisation.

UK research on cooperative learning

Considering the extent to which "cooperative learning" has been endorsed by central agencies in the UK, albeit in an unspecified manner, it is surprising how little research into British practice has been undertaken. Prominent in this field have been Cowie and Rudduck (eg. 1990), who argue that there is a political justification for the development of cooperative skills and a desire from industry to have employees who can work effectively in teams. They note that much UK work is qualitative and descriptive in character, being as concerned with the process as with the product (eg. Stenhouse, 1970; Barnes, 1976; Salmon & Claire, 1984). Cowie and Rudduck (1990) remark that however strong the teacher's belief in the value of cooperative learning, the pupils may be unconvinced and express preference for traditional methods with which they feel more comfortable.

Elsewhere in the UK, Foot and his co-workers have sought to articulate some of the key theoretical issues in cooperative learning (Foot *et al.*, 1990) and have conducted empirical studies designed to elucidate the impact of various organisational parameters in cooperative learning structures. Cooperative learning activities are seen as the ideal setting for cognitive

conflict and conflict resolution in a Piagetian sense. It is observed that workers such as Russell (1981) and Light *et al.* (1987) found no advantage for dyadic performance over solitary performance where purportedly collaborative tasks were unstructured, since most of the decision-making was done by the more dominant child of the pair and the other merely complied.

During cooperative interaction it is posited that children have the opportunity for planning strategies, verifying ideas and encountering the symbolic representation of intellectual acts through peer communication. The ineffectiveness of an individual in the cooperative learning situation may reflect failure to distribute limited cognitive resources adequately between the various complex components of the task. When the task is perceived to be too complicated, the child may focus upon immediate task demands while neglecting wider or more long-term aspects. In this latter context, the signalling and detection of non-understanding may be crucial, and therefore a feature of cooperative interaction which should be articulated or structured in initiating and monitoring such procedures. Foot *et al.* (1990) also emphasised the importance of the child's perception of their own role and of learning and teaching strategies in general, the sensitivity of cooperative learners to the needs of others and the need for effective communication.

In an empirical study, Foot *et al.* (1988) found the amount of participation in the learning task influenced memory test scores of tutees, thereby highlighting the need for interaction to be structured to eliminate tutee passivity.

Foot and Barron (1990) then explored the impact of pre-existing friendship patterns on the process and outcome of cooperative learning. Although allocation to groups on the basis of pre-existing friendships might be expected to result in children focussing their resources on the informational components of the task, since the burden of managing the social demands should be minimised, in fact the opposite was found. Good friends indulged in their relationship and learnt no more effectively than non-friends. The authors noted that previous research on the effectiveness of allocation by friendship was somewhat conflicting. An incentive condition was also included since again previous research was conflicting – Garbarino (1975) had found that incentives had a deleterious effect in a cross-age tutoring experiment with 12- and 7-year-olds. The incentive condition had no impact on test scores. However, the authors felt that these findings might not be generalisable since differentiation of task type and complexity could yield different results.

Two studies exploring the relative impact of two types of training were reported by Barron and Foot (1991). "Non-elaborated" training covered only the procedures necessary for carrying out the task, while "elaborated" training also covered the generalisable principles involved. Eight-year-old children in cooperative pairs performed better on spatial-numeric and item recall tasks when general principles were specifically taught, in contrast to a condition where they were encouraged to educe such principles themselves without guidance.

Crucial issues

There is clearly considerable variety and conflict in the literature, both within and between continents. To what extent is competition between working groups a necessary and desirable feature of the organisation of cooperative learning, if achievement gains are to be as good or better than other pedagogical procedures? Is some system of external accountability of the individual within a cooperative working group also necessary, or can activities be organised so that this function is carried out within the group itself? Is there really a need for extrinsic reinforcement, or does this vary according to societal and cultural expectations, and may its inappropriate insertion actually result in worse outcomes?

Above all, perhaps the most striking disparity between the Northern American and European approaches concerns the degree of detailed structure injected by the supervising professional. Some North American cooperative learning structures appear extremely rigid, rule-bound and prescriptive, while at the other end of the dimension some British approaches, while warm, fuzzy and comfortable, demonstrate an organisational looseness bordering on chaos. Perhaps the crucial issue then, particularly for the practising teacher, is how much structure and of what kind to engineer into any proposed cooperative learning activity, taking into account the age and ability of the pupils, the length and complexity of the task, and the degree of inherent structure and differentiation in the curriculum materials involved. To what extent should cooperative learning be "scaffolded", and how?

It might be that the special case of a cooperative learning format known as peer tutoring, involving dyadic rather than larger group interaction, offers practitioners a useful tool for exploring these issues in a controlled manner.

Peer tutoring – definitions

A traditional definition of peer tutoring might have been: "more able pupils helping less able pupils in cooperative working pairs carefully organised by a teacher". The differential in ability between the more competent tutor and the less competent tutee assumed by this definition was sometimes taken to imply a differential in age also. In fact, in recent times cross-age tutoring has been widely supplemented and sometimes substituted by same-age tutoring. Furthermore, new organisational structures for dyadic tutoring sometimes dispense with the need for an ability differential.

Peer tutoring is certainly a very ancient practice (Topping, 1988), but it should not be dismissed as merely a re-hash of the old "monitorial" system, since recent research and development have refined and extended the method considerably.

Initial objections to peer tutoring often embody resistance to more able pupils being "used" to help less able pupils. Peer tutoring is indeed difficult

to justify if this is the case, but properly organised peer tutoring targets achievement gains by both tutors and tutees – and often delivers exactly this. For the tutor, the experience should be "Learning by Teaching".

For teachers new to cooperative learning, pairs rather than small groups might prove easier to organise effectively, roles might be more readily specifiable thus reducing complexity for the pupils, and it might be easier for the teacher to see where the structure is going wrong for those dyads where this occurs.

Goodlad (1979) proposed a host of reasons why peer tutoring might work. It might make the tutors feel more adequate and responsible, lead them to make more meaningful use of their knowledge and information, lead them to revise and reinforce their knowledge of fundamentals, improve their self-image if they felt accorded the status of teachers and give them insight into the teaching process causing them to respond better to their own professional teachers. The tutees might benefit from individualised one-to-one instruction, receive more teaching and experience more practice while staying on task for greater proportions of contact time, receive more immediate and frequent feedback, error correction and reinforcement, and might also enjoy the companionship, possibly responding on a personal level better to peers of their own age and cultural background than to a teacher possessing neither of these qualities. The benefits for the teachers might be that the promotion of achievement would proceed at least equally effectively while relieving the teacher of the stress of monitoring a large group simultaneously, freeing the teacher from routine tasks and enabling them to conduct more precise, technical and sophisticated instruction and monitoring.

Research on peer tutoring

A number of substantial reviews of research on peer tutoring have been published during the last 20 years, far too many to mention here. The volume *Children Teach Children* published in 1971 by Gartner *et al*. was the first of significance, but Allen's (1976) *Children As Teachers* reached a much wider audience and remains a classic. Allen (1976) noted that more studies reported positive effects for tutors than for tutees. Peer tutoring had proved effective across barriers of gender, race and social class.

Subsequently, Sharpley and Sharpley (1981) reviewed 82 peer tutor studies, concluding that same-age tutors were as effective as cross-age tutors in inducing cognitive advances in tutees, and also that same-age tutors were themselves more likely to derive cognitive benefits as a result of their tutoring experiences. Most peer tutoring studies focussed on reading skills, but 16 dealt with mathematics and a further four with spelling and language skills, while other curriculum areas included social science, French German and Spanish languages, written expression, creative thinking, problem solving, drugs and sexuality and birth control. Results on the effects of extrinsic

reinforcement were equivocal, three studies claiming that tangible reinforcement improved functioning at least on simple tasks in the short term, while another three found no such effect. Studies of the relative effectiveness of trained versus untrained tutors indicated that although unstructured programmes using untrained tutors could succeed, they were likely to have a lower success rate. However, of the 82 studies, in terms of academic outcomes for tutors, 21 studies reported positive effects, none reported negative effects, but 29 reported non-significant effects. For academic outcomes for tutees, 35 studies reported positive effects, no studies reported negative effects, but 27 reported non-significant effects.

Moving on from the traditional literature review, Cohen *et al.* (1982) conducted a meta-analysis of 65 studies of peer tutoring. In 45 of 52 studies, tutored students out-performed control students, while in 6 studies control students did better and in one study there was no difference. The average effect size overall was 0.4. Studies involving tutor training and structured tutoring tended to produce larger effect sizes. Eight studies reported improvements in student attitudes and seven studies reported improvements in tutee self concept although this latter effect size was small. Cohen *et al.* (1982) concluded "these programmes have definite and positive effects on the academic performance and attitudes of those who receive tutoring and also have positive effects on children who serve as tutors". However, "tutoring programmes have much smaller effects on the self concept of children", despite the "anecdotal reports of dramatic changes" in the literature.

More recently, Topping (1988) produced a "review of reviews", commenting on the increasing involvement as tutors and tutees of students who have special needs, and the extension of peer tutoring methodology to tutors as young as three years and to increasingly exotic curriculum areas. The literature was also reviewed by Goodlad and Hirst in 1989. A further volume edited by Goodlad and Hirst (1990) contained reports of many extremely diverse studies in the UK and elsewhere. In this volume, Fitz-Gibbon (1990) reported five controlled studies, of which the two involving same-age tutoring were considered unsuccessful, leading her to conclude that cross-age tutoring may have more promise, although this is not supported by the weight of meta-analytic evidence cited earlier. Both of the Goodlad and Hirst volumes contain considerable reportage of peer tutoring in higher education, particularly Goodlad's own work in involving undergraduates to help tutor science, mathematics and engineering in high schools.

Increasingly, specific structured "packages" for peer tutoring of basic skills have been developed, researched and disseminated. Wheldall and Colmar (1990) have reviewed evidence on the effectiveness of the "Pause, Prompt, Praise" technique for reading, while Topping (1990) and Topping and Lindsay (1992) have done the same for the "Paired Reading" technique. Topping and Whiteley (1990) reported that using the Paired Reading technique children tutored by parents did no better in the short term than children

tutored by peers, with effect sizes of the order of 0.87 for reading accuracy and 0.77 for reading comprehension. Peer tutor gains were greater than peer tutee gains in reading accuracy, but this difference did not reach statistical significance. Subjective feedback from same-age peer tutees was more positive than that from cross-age peer tutees, although tutors tended to be equally positive irrespective of age differential.

Peer tutoring continues to spread across the curriculum, into chemistry (Bland & Harris, 1988), social studies (Maheady *et al*., 1988), foreign language teaching (Fitz-Gibbon & Reay, 1982), spelling (Oxley & Topping, 1990) and higher order reading skills (Moore, 1988).

Reciprocal peer tutoring

As interest has focussed on the potency of the act of tutoring as a vehicle for learning, the assumption that tutors should necessarily be more able than the tutees, or at least at a level of greater mastery of the curriculum material involved, has begun to be challenged. Workers have increasingly experimented with reciprocation of tutor-tutee roles within dyadic pairings. More ethically acceptable as incorporating more immediately equal opportunities, this format appears also to offer participants greater variety and novelty of learning experiences. However, where reciprocal tutoring involves zero ability differentials in the pair, some continuously available external reference point against which their efforts can be checked for accuracy becomes essential, (eg. self-correcting curriculum materials or activities).

Thus Pigott *et al*. (1986) reported on reciprocal peer tutoring in teams of four elementary school children working on routine arithmetic drill. Underachieving pupils improved their performance to a level indistinguishable from that of their classmates and these gains were maintained at follow-up. Palincsar and Brown (1986) similarly described a procedure for developing reading comprehension skills in which specific task roles rotated round a small group: predicting, question generating, summarising and clarifying. Evaluation results were impressive. Dyadic reciprocal peer tutoring involving peer-managed group contingencies was reported by Fantuzzo *et al*. (1990), who found consistent increases in rate of accurate arithmetic performance to a level significantly above the rates of untreated controls, which was maintained at follow up. In the UK, Brierley *et al*. (1989) described dyadic class-wide reciprocal peer tutoring of spelling on a large scale. Results were suggestive of generalised improvement in spelling skills.

The effectiveness of reciprocal peer tutoring would appear to highlight the importance of task-focussed structured interaction in cooperative learning pairs or groups. This is very different from a traditional conception of peer tutoring, with its associated preoccupation with the "qualities" of a "good" tutor. In reciprocal peer tutoring, if the process is properly structured, the input and output (product) stages might take care of themselves. Nor is

the RPT format any less a way of "structuring positive interdependence", since within the structure neither member of the pair can function without the other, and the motivation of tutors to be positive towards their tutees could be heightened by the knowledge that they themselves will become tutees during the following session.

Special needs and ethnic minorities

Turning now to consider again all forms of cooperative learning, including the special case of peer tutoring, it is noteworthy that the literature on the effectiveness of these approaches with pupils with special educational needs and pupils from ethnic minorities is substantial.

Many studies have deployed under-achieving pupils as tutors in order to improve the attainment and self-esteem of the tutors. An extreme example is reported by Custer and Osguthorpe (1983), who arranged for mentally handicapped pupils to tutor their non-handicapped peers in sign language, with the result that the sign language competence of both tutors and tutees improved and social interaction between the two groups improved even more. The deployment as tutors of students with emotional and behavioural difficulties has also been widely explored.

A classic study of that Maher (1984), who deployed high school behaviour difficulty students as cross-age tutors for elementary school pupils with learning difficulties in reading, language and mathematics. Gains in attainment for both tutors and tutees are detailed, and disciplinary referrals of tutors fell from an average of six during the baseline period to two during the intervention period, stabilising at one during follow up.

Cook *et al.* (1986) carried out a meta-analysis of studies of special needs students as tutors of others. Over 19 studies, involvement in tutoring raised the performance of tutors and tutees as compared to that of controls, with tutees usually achieving greater gains than tutors. Behaviour ratings of tutors showed some improvement, more than those of tutees, and both tutors and tutees showed improvement in attitude towards school and/or the curriculum area of tutoring. However, changes in measures of self-image and sociometric integration tended not to be statistically significant.

Osguthorpe and Scruggs (1990) reviewed 26 studies deploying special education students as tutors, and noted that 23 found the tutors and/or tutees performed better on outcome measures. Again, attempts to demonstrate the effectiveness of tutoring on "self-esteem" had generally been unsuccessful, possibly owing to problems of measurement. These authors also emphasised the importance of adequate training and supervision. A wider ranging review of cooperative learning among special students (Ashman & Elkins, 1990) also concluded that structured interaction could lead to significant gains, but in unstructured conditions many of the moves tended to be made by one student without consultation or collaboration with peers.

The potential value of peer tutoring as a possible method for multi-ethnic education was raised by Fitz-Gibbon in 1983. In 1985 Slavin reviewed research on the use of his cooperative learning methods in multi-ethnic classrooms, and claimed consistently positive effects on inter-group relations, as well as on the achievement of minority and majority students. Sixteen of the 19 studies reviewed demonstrated improvement in some aspect of friendship between students of different ethnicities.

More recently, Cowie (1990) and her associates have explored the effectiveness of an enhanced cooperative group work curriculum in multi-ethnic classrooms in improving social relationships. Methods used by participating teachers included trust-building exercises, problem-solving groups, role-playing, discussion groups, report back sessions and debriefing. After the intervention experimental children tended to like each other more, irrespective of race and gender, and there was also evidence of increased cross-race and cross-gender preference choices, coupled with a reduction in negative stereotyping of other ethnic groups, as compared to children in control classes (Smith *et al*., 1989).

A review of cooperative learning in multi-ethnic classrooms by Sharan (1990) mainly covered studies adopting a Slavin-type methodology but did take an international perspective. Noting that the positive findings concerning the effects of cooperative learning on ethnic relations reported in a meta-analysis by Johnson *et al*. (1983) had subsequently been widely replicated, Sharan nevertheless commented that little evidence was yet available concerning generalisation to relationships outside school, or indeed even to other pupils within school who had not participated in the cooperative learning experience. Sharan (1990) also referred to the meta-analysis of 21 co-operative learning studies in multi-ethnic classrooms by Miller and Davidson-Podgorny (1987), who found inter-group relations particularly improved by task structures involving interdependence, random task role assignment and equal representation of the various ethnic groups.

Conclusions and issues

Although the generic label "cooperative learning" has been used to cover a multitude of educational activities, the very substantial research literature suggests that the approach has great promise in terms of yielding cognitive, social and affective gains. Special needs students and ethnic minority pupils can benefit in terms of improved social integration as well as attainment. The indications are that the cost-effectiveness of the method can compare well to that of more traditional forms of instruction. Practice in the United Kingdom has tended to lack the clear structure characteristic of North American work, and objectives, procedures and evaluation would benefit from sharper definition. The subspecies of peer tutoring has frequently shown attainment and attitude gains for both tutors and tutees.

Further research is certainly required. Comparative research on the relative effectiveness in different situations of a variety of organisational structures must continue. There is a great need for further (unfortunately time-consuming) intensive study of the process of cooperative learning – who actually *does* what, with which and to whom, and how this relates to objectives, training and assumptions. Where combined treatments are applied, research to disentangle the effect of component parts must be attempted as a first step towards elucidating aptitude × treatment interactions.

Two further major issues arise. Given a larger number of teachers with a repertoire of various cooperative learning strategies, explorations may commence into the impact of offering students consultation or informed choice regarding preferred methods – thereby hopefully raising their metacognitive awareness. A more practical and pressing issue concerns the large-scale replicative durability of cooperative learning approaches. At a time when the UK education system is stressed, confused and starved of resources, with in-service training and support systems crumbling, if cooperative learning is not easy to do then it may not be done at all. Thus methods need to be tailored and tempered to fit as easily and naturally as possible within a multitude of classrooms – not just those of teachers who are well-motivated and well-organised enough to be able to attend in-service courses. Further development of class-wide (eg. Greenwood *et al*., 1987), same-age and reciprocal peer tutoring may be anticipated in this context.

It will be important that teachers see cooperative learning as an alternative and effective route to National Curriculum attainment targets, which can have the additional benefit of positive attitudinal and social side effects. Furthermore, such activities promote the pursuance of the cross-curricular theme of "Education for Citizenship". Cooperative learning certainly *has* a place in the individualistic and competitive world of today – in fact, it is needed all the more.

Resources

The magazine *Co-operative Learning* incorporated an excellent resource guide in volume 11, number 1, September 1990. The magazine is published under the auspices of the "International Association for the Study of Co-operation in Education", Box 1582, Santa Cruz, California 95061-1582, USA. In the UK, a series of four books and training guide called *Learning Together, Working Together*, by Helen Cowie and Jean Rudduck, is available from BP Educational Service, PO Box 30, Blacknest Road, Blacknest, Alton, Hampshire GU34 4PX. BP also sponsor a resource pack for HE into schools tutoring. A useful workbook on cooperative grouping is: *Talking and Learning in Groups* by N. Bennett and E. Dunne, London, Macmillan, 1990. Stewart Ehly and Stephen Larsen incorporated a number of useful materials in their book *Peer Tutoring for Individualised Instruction*, published in 1980 in Boston by Allyn and Bacon. This was followed by a simpler manual (*Peer Tutoring: A Guide for School Psychologists*), written by Ehly and published by the National Association

of School Psychologists (in the USA) in Kent, Ohio (ISBN 0-932955-01-0). There is an associated video tape. A useful book for teachers is *Peer Teaching and Collaborative Learning in the Language Arts* by Elizabeth McAllister (1990), available from the ERIC Clearinghouse on Reading and Communication Skills at Indiana University, Bloomington. The book by Goodlad and Hirst (1989) includes a useful planning guide for organising peer tutoring (see reference). In the USA, David and Roger Johnson edited a volume entitled *Structuring Co-operative Learning: Lesson Plans for Teachers*, published in 1984 in New Brighton Minnesota by the Interaction Book Company. Robert Slavin has been extremely prolific, but the most up-to-date and practical handbook covering his work was published in 1990 (see references). Returning to the UK, the Cleveland Learning Support Service has produced a useful instructional video on peer tutoring in conjuction with Teesside Polytechnic. Keith Topping has produced a practical handbook for teachers entitled *The Peer Tutoring Handbook: Promoting Co-operative Learning*, published in 1988 in London by Croom Helm and in Cambridge Massachusets by Brookline. The Kirklees Paired Learning Project has produced a number of video training packs on peer tutoring in basic skills and published *Paired Learning* (The Oastler Centre, 103 New Street, Huddersfield HD1 2UA West Yorkshire). More recent resources are available from the Centre for Paired Learning, University of Dundee, DD1 4HN. The Peer Tutoring Consortium facilitates networking, produces newsletters and maintains registers of training providers and resource materials (contact at the Curriculum, Evaluation and Management Centre, School of Education, University of Newcastle upon Tyne, NE1 7RU).

References

Allen, V.L. (ed.) (1976). *Children as Teachers: Theory and Research on Tutoring*. New York: Academic Press

Ashman, A.F. & Elkins, J. (1990). Co-operative Learning Among Special Students. In Foot, H.C. *et al*. (1990a)

Barnes, D. (1976). *From Communication to Curriculum*. Harmondsworth: Penguin

Barron, A.-M. & Foot, H.C. (1991). Peer tutoring and tutor training. *Educational Research*, **33**, 3, 174–185

Bennett, N. (1987). Co-operative learning: Children do it in groups – or do they? *Educational and Child Psychology*, **4**, 3 & 4, 7–18

Bennett, S.N., Desforges, C.W., Cockburn, A. & Wilkinson, B. (1984). *The Quality of Pupil Learning Experiences*. London: Lawrence Erlbaum

Bland, M. & Harris, G. (1988). Peer tutoring as part of colloborative teaching in chemistry. *Support for Learning*, **3**, 4, 215–8

Brierley, M., Hutchinson, P., Topping, K. & Walker, C. (1989). Reciprocal peer tutored cued spelling with ten year olds. *Paired Learning*, **5**, 136–40

Cohen, P.A., Kulik, J.A. & Kulik, C.-L.C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, **19**, 2, 237–48

Cook, S.B., Scruggs, T.E., Mastropieri, M.A. & Casto, G.C. (1986). Handicapped students as tutors. *Journal of Special Education*, **19**, 4, 483–92

Cotton, J. & Cook, M. (1982). Meta-analyses and the effect of various systems: Some different conclusions from Johnson *et al.*. *Psychological Bulletin*, **92**, 176–83

Cowie, H. (1990). Ethnic relations in middle schools. Paper delivered at IASCE 5th International Convention on Co-operative Learning, Baltimore Md., July 1990

Cowie, H. & Rudduck, J. (1988). Testing teams: Is GCSE really encouraging more group work? *The Times Educational Supplement*, 15.4.88, page 21.

Cowie, H. & Rudduck, J. (1990). Learning from one another: The challenge. In Foot, H.C. *et al*. (1990a)

Custer, J.D. & Osguthorpe, R.T. (1983). Improving social acceptance by training handicapped students to tutor their non-handicapped peers. *Exceptional Children*, **50**, 2, 173–5

Davidson, N. (1990). Co-operative learning research in mathematics. Paper delivered at IASCE 5th International Convention on co-operative learning. Baltimore Md., July 1990

Doyle, W. (1986). Classroom organisation and management. In: Wittrock, M.C. (ed.). *Handbook of Research on Teaching* (3rd edition) New York: Macmillan

Fantuzzo, J.W., Polite, K. & Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *Journal of School Psychology*, **28**, 4, 309–23

Fitz-Gibbon, C.T. (1983). Peer tutoring: A possible method for multi-ethnic education. *New Community*, **11**, 160–6

Fitz-Gibbon, C.T. (1990). Success and failure in peer tutoring experiments. In: Goodlad, S. & Hirst, B. (eds) (1990)

Fitz-Gibbon, C.T. & Reay, D.G. (1982). Peer tutoring: Brightening up foreign language teaching in an urban comprehensive school. *British Journal of Language Teaching*, **20**, 1, 39–46

Foot, H. & Barron, A.-M. (1990). Friendship and task management in children's peer tutoring. *Educational Studies*, **16**, 3, 237–50

Foot, H.C., Shute, R.H. & Morgan, M.J. (1988). Peer tutoring and children's memory. In: Gruneberg, M.M., Morris, P.E. & Sykes, R.N. (eds). *Practical Aspects of Memory: Current Research and Issues*. Chichester: Wiley

Foot, H.C., Shute, R.H., Morgan, M.J. & Barron, A.-M. (1990). Theoretical issues in peer tutoring. In Foot, H.C. *et al*. (1990a)

Foot, H.C., Morgan, M.J. & Shute, R.J. (eds) (1990a). *Children Helping Children*. Chichester. Wiley

Garbarino, J. (1975). The impact of anticipated reward upon cross-age tutoring. *Journal of Personality and Social Psychology*, **32**, 421–8

Gartner, S., Kohler, M. & Riessman, F. (1971). *Children Teach Children: Learning by Teaching*. New York: Harper and Row

Goodlad, S. (1979). *Learning by Teaching: An Introduction to Tutoring*. London: Community Service Volunteers

Goodlad, S. & Hirst, B. (1989). *Peer Tutoring: A Guide to Learning by Teaching*. London: Kogan Page; New York: Nichols

Goodlad, S. & Hirst, B. (eds) (1990). *Explorations in Peer Tutoring*. Oxford: Blackwell

Greenwood, C.R., Carta, J.J. & Kamps, D. (1990). Teacher-mediated versus peer-mediated instruction: A review of educational advantages and disadvantages. In Foot, H.C. *et al*. (1990a)

Greenwood, C.R., Dinwiddie, G., Bailey, V., Carta, J.J., Dorsey, D., Kohler, F.W., Nelson, C., Rotholz, D. & Schulte, D. (1987). Field replication of classwide peer tutoring. *Journal of Applied Behavior Analysis*, **20**, 2, 151–60

Her Majesty's Inspectorate (1978). *Primary Education in England*. London: HMSO
Johnson, D.W. & Johnson, R.T. (1989) *Co-operation and Competition: Theory and Research*. Edina, MN: Interaction Book Co.
Johnson, D.W. & Johnson, R.T. (1991). *Learning Together and Alone* (3rd edition). Englewood Cliffs, New Jersey: Prentice Hall
Johnson, D., Johnson, R. & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and meta-analysis of the research. *Review of Educational Research*, **68**, 446–52
Johnson, D.W., Maruyama, G., Johnson, R., Nelson, D. & Skon, L. (1981). Effects of co-operative, competitive and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, **89**, 47–62
Johnson, R.T. & Johnson, D.W. (1983). Effects of co-operative, competitive and individualistic learning experiences on social development. *Exceptional Children*, **49**, 4, 323–9
Levin, H.M. & Meister, G. (1986). Is CAI cost-effective? *Phi Delta Kappan*, **67**, 745–9
Light, P., Foot, T., Colburn, C. & McLelland, I. (1987). Collaborative interactions at the microcomputer keyboard. *Educational Psychology*, **7**, 13–21
MacAulay, D.J. (1990). Classroom environment: a literature review. *Educational Psychology*, **10**, 3, 239–53
McGlynn, R. (1982). A comment on the meta-analysis of goal structures. *Psychological Bulletin*, **92**, 184–5
Maheady, L., Sacca, M.K. & Harper, G.F. (1988). Classwide peer tutoring with mildly handicapped high school students. *Exceptional Children*, **55**, 1, 52–9
Maher, C.A. (1984). Handicapped adolescents as cross-age tutors: Program description and evaluation. *Exceptional Children*, **51**, 1, 56–63
Miller, N. & Davidson-Podgorny, G. (1987). Theoretical models of intergroup relations and the use of co-operative teams as an intervention for desegregated settings. In: Hendrick, C. (ed.). *Group Processes and Intergroup Relations*. Newbury Park, California: Sage.
Moore, P.J. (1988). Reciprocal teaching and reading comprehension: A review. *Journal of Research in Reading*, **11**, 1, 3–14
Osguthorpe, R.T. & Scruggs, T.E. (1990). Special education students as tutors: A review and analysis. In: Goodlad, S. & Hirst, B. (eds) (1990)
Oxley, L. & Topping, K. (1990). Peer tutored cued spelling with seven to nine year-olds. *British Educational Research Journal*, **16**, 1, 63–78
Palincsar, A.S. & Brown, A.L. (1986). Interactive teaching to promote independent learning from text. *The Reading Teacher*, **39**, 8, 771–7
Pigott, H.E., Fantuzzo, J.W. & Clement, P.W. (1986). The effects of reciprocal peer tutoring and group contingencies on the academic performance of elementary school children. *Journal of Applied Behavior Analysis*, **19**, 1, 93–8
Plowden Report (1967). *Children and their Primary Schools*. (Report of the Central Advisory Council for Education: England). London: HMSO
Rudduck, J. & Cowie, H. (1988). An introduction to co-operative group work. *Educational and Child Psychology*, **5**, 4, 91–102
Russell, J. (1981). Dyadic interaction in a logical reasoning problem requiring inclusion ability. *Child Development*, **52**, 1322–5

Russell, T. & Ford, D.F. (1983). Effectiveness of peer tutors vs. resource teachers. *Psychology in the Schools*, **20**, 436–41

Salmon, P. & Claire, H. (1984). *Classroom Collaboration*. London: Routledge & Kegan Paul

Sharan, S. (1990). Co-operative learning and helping behaviour in the multi-ethnic classroom. In Foot, H.C. *et al*. (1990a)

Sharpley, A.M. & Sharpley, C.F. (1981). Peer tutoring: A review of the literature. *Collected Original Resources in Education*, **5**, 3, 7-C11

Sherman, J.A. & Harris, V.W. (1975). Effects of peer tutoring and homework assignments on classroom performance. In: Thompson, T. & Dockens, W.S. (eds) *Applications of Behavior Modification*. New York: Academic Press

Slavin, R.E. (1985). Co-operative learning: Applying contact theory in desegregated schools. *Journal of Social Issues*, **41**, 3, 45–62

Slavin, R.E. (1990). *Co-operative Learning: Theory, Research and Practice*. Englewood Cliffs, New Jersey: Prentice Hall

Smith, P.K., Boulton, M. & Cowie, H. (1989). Ethnic relations in middle school. *ESRC Progress Report*. Psychology Department, University of Sheffield

Stenhouse, L. (1970). *The Humanities Project: An Introduction*. London: Heinemann

Tizard, J., Schofield, W.N. & Hewison, J. (1982). Collaboration between teachers and parents in assisting children's reading. *British Journal of Educational Psychology*, **52**, 1–15

Topping, K. (1988). *The Peer Tutoring Handbook: Promoting Co-operative Learning*. London: Croom Helm; Cambridge, MA: Brookline Books

Topping, K. (1990). Peer tutored paired reading: Outcome data from ten projects. In: Goodlad, S. & Hirst, B. (eds) (1990)

Topping, K. & Lindsay, G.A. (1992). Paired reading: A review of the literature. *Research Papers in Education*, **7**, 3

Topping, K. & Whiteley, M. (1990). Participant evaluation of parent-tutored and peer-tutored projects in reading. *Educational Research*, **32**, 1, 14–32

Wheldall, K. & Colmar, S. (1990). Peer tutoring for low-progress readers using "Pause Prompt and Praise". In Foot, H.C. *et al*. (1990a)

Wheldall, K., Morris, M., Vaughan, P. & Yin Yuk, N.G. (1981). Rows vs. Tables: An example of the use of behavioural ecology in two classes of eleven year old children. *Educational Psychology*, **1**, 2, 171–83

Part XIX

MORAL EDUCATION

88

THE COGNITIVE–DEVELOPMENTAL APPROACH TO MORAL EDUCATION

L. Kohlberg

In this article, I present an overview of the cognitive-developmental approach to moral education and its research foundations, compare it with other approaches, and report the experimental work my colleagues and I are doing to apply the approach.

I. Moral stages

The cognitive-developmental approach was fully stated for the first time by John Dewey. The approach is called *cognitive* because it recognizes that moral education, like intellectual education, has its basis in stimulating the *active thinking* of the child about moral issues and decisions. It is called developmental because it sees the aims of moral education as movement through moral stages. According to Dewey:

> The aim of education is growth or *development*, both intellectual and moral. Ethical and psychological principles can aid the school in the *greatest of all constructions – the building of a free and powerful character*. Only knowledge of the *order and connection of the stages in psychological development can insure this*. Education is the work of *supplying the conditions* which will enable the psychological functions to mature in the freest and fullest manner.[1]

Dewey postulated three levels of moral development: 1) the *pre-moral* or *preconventional* level "of behavior motivated by biological and social impulses with results for morals," 2) the *conventional* level of behavior "in which the individual accepts with little critical reflection the standards of his group," and 3) the *autonomous* level of behavior in which "conduct is guided by the individual thinking and judging for himself whether a purpose is good, and does not accept the standard of his group without reflection."[a]

Source: *Phi Delta Kappan*, 1975, 56(10), 670–677.

Table 1 Definition of moral stages

I. Preconventional level

At this level, the child is responsive to cultural rules and labels of good and bad, right or wrong, but interprets these labels either in terms of the physical or the hedonistic consequences of action (punishment, reward, exchange of favors) or in terms of the physical power of those who enunciate the rules and labels. The level is divided into the following two stages:

Stage 1: *The punishment-and-obedience orientation*. The physical consequences of action determine its goodness or badness, regardless of the human meaning or value of these consequences. Avoidance of punishment and unquestioning deference to power are valued in their own right, not in terms of respect for an underlying moral order supported by punishment and authority (the latter being Stage 4).

Stage 2: *The instrumental-relativist orientation*. Right action consists of that which instrumentally satisfies one's own needs and occasionally the needs of others. Human relations are viewed in terms like those of the marketplace. Elements of fairness, of reciprocity, and of equal sharing are present, but they are always interpreted in a physical, pragmatic way. Reciprocity is a matter of "you scratch my back and I'll scratch yours," not of loyalty, gratitude, or justice.

II. Conventional level

At this level, maintaining the expectations of the individual's family, group, or nation is perceived as valuable in its own right, regardless of immediate and obvious consequences. The attitude is not only one of *conformity* to personal expectations and social order, but of loyalty to it, of actively *maintaining*, supporting, and justifying the order, and of identifying with the persons or group involved in it. At this level, there are the following two stages:

Stage 3: *The interpersonal concordance or "good boy – nice girl" orientation*. Good behavior is that which pleases or helps others and is approved by them. There is much conformity to stereotypical images of what is majority or "natural" behavior. Behavior is frequently judged by intention – "he means well" becomes important for the first time. One earns approval by being "nice."

Stage 4: *The "law and order" orientation*. There is orientation toward authority, fixed rules, and the maintenance of the social order. Right behavior consists of doing one's duty, showing respect for authority, and maintaining the given social order for its own sake.

III. Postconventional, autonomous, or principled level

At this level, there is a clear effort to define moral values and principles that have validity and application apart from the authority of the groups or persons holding these principles and apart from the individual's own identification with these groups. This level also has two stages:

Stage 5: *The social-contract, legalistic orientation*, generally with utilitarian overtones. Right action tends to be defined in terms of general individual rights and standards which have been critically examined and agreed upon by the whole society. There is a clear awareness of the relativism of personal values and opinions and a corresponding emphasis upon procedural rules for reaching consensus. Aside from

what is constitutionally and democratically agreed upon, the right is a matter of personal "values" and "opinion." The result is an emphasis upon the "legal point of view," but with an emphasis upon the possibility of changing law in terms of rational considerations of social utility (rather than freezing it in terms of Stage 4 "law and order"). Outside the legal realm, free agreement and contract is the binding element of obligation. This is the "official" morality of the American government and constitution.

Stage 6: *The universal-ethical-principle orientation*. Right is defined by the decision of conscience in accord with self-chosen *ethical principles* appealing to logical comprehensiveness, universality, and consistency. These principles are abstract and ethical (the Golden Rule, the categorical imperative); they are not concrete moral rules like the Ten Commandments. At heart, these are universal principles of *justice*, of the *reciprocity* and *equality* of human *rights*, and of respect for the dignity of human beings as *individual persons* ("From Is to Ought," pp. 164, 165).

– Reprinted from *The Journal of Philosophy*, October 25, 1973

Dewey's thinking about moral stages was theoretical. Building upon his prior studies of cognitive stages, Jean Piaget made the first effort to define stages of moral reasoning in children through actual interviews and through observations of children (in games with rules).[2] Using this interview material, Piaget defined the pre-moral, the conventional, and the autonomous levels as follows: 1) the *pre-moral stage*, where there was no sense of obligation to rules; 2) the *heteronomous stage*, where the right was literal obedience to rules and an equation of obligation with submission to power and punishment (roughly ages 4–8); and 3) the *autonomous stage*, where the purpose and consequences of following rules are considered and obligation is based on reciprocity and exchange (roughly ages 8–12).[b]

In 1955 I started to redefine and validate (through longitudinal and cross-cultural study) the Dewey-Piaget levels and stages. The resulting stages are presented in Table 1.

We claim to have validated the stages defined in Table 1. The notion that stages can be *validated* by longitudinal study implies that stages have definite empirical characteristics.[3] The concept of stages (as used by Piaget and myself) implies the following characteristics:

1. Stages are "structured wholes," or organized systems of thought. Individuals are *consistent* in level of moral judgment.
2. Stages form an *invariant sequence*. Under all conditions except extreme trauma, movement is always forward, never backward. Individuals never skip stages; movement is always to the next stage up.
3. Stages are "hierarchical integrations." Thinking at a higher stage includes or comprehends within it lower-stage thinking. There is a tendency to function at or prefer the highest stage available.

Each of these characteristics has been demonstrated for moral stages. Stages are defined by responses to a set of verbal moral dilemmas classified according to an elaborate scoring scheme. Validating studies include:

1. A 20-year study of 50 Chicago-area boys, middle- and working-class. Initially interviewed at ages 10–16, they have been reinterviewed at three-year intervals thereafter.
2. A small, six-year longitudinal study of Turkish village and city boys of the same age.
3. A variety of other cross-sectional studies in Canada, Britain, Israel, Taiwan, Yucatan, Honduras, and India.

With regard to the structured whole or consistency criterion, we have found that more than 50% of an individual's thinking is always at one stage, with the remainder at the next adjacent stage (which he is leaving or which he is moving into).

With regard to invariant sequence, our longitudinal results have been presented in the *American Journal of Orthopsychiatry* (see footnote 8), and indicate that on every retest individuals were either at the same stage as three years earlier or had moved up. This was true in Turkey as well as in the United States.

With regard to the hierarchical integration criterion, it has been demonstrated that adolescents exposed to written statements at each of the six stages comprehend or correctly put in their own words all statements at or below their own stage but fail to comprehend any statements more than one stage above their own.[4] Some individuals comprehend the next stage above their own; some do not. Adolescents prefer (or rank as best) the highest stage they can comprehend.

To understand moral stages, it is important to clarify their relations to stage of logic or intelligence, on the one hand, and to moral behavior on the other. Maturity of moral judgment is not highly correlated with IQ or verbal intelligence (correlations are only in the 30s, accounting for 10% of the variance). Cognitive development, in the stage sense, however, is more important for moral development than such correlations suggest. Piaget has found that after the child learns to speak there are three major stages of reasoning: the intuitive, the concrete operational, and the formal operational. At around age 7, the child enters the stage of concrete logical thought: He can make logical inferences, classify, and handle quantitative relations about concrete things. In adolescence individuals usually enter the stage of formal operations. At this stage they can reason abstractly, i.e., consider all possibilities, form hypotheses, deduce implications from hypotheses, and test them against reality.[c]

Since moral reasoning clearly is reasoning, advanced moral reasoning depends upon advanced logical reasoning; a person's logical stage puts a

certain ceiling on the moral stage he can attain. A person whose logical stage is only concrete operational is limited to the preconventional moral stages (Stages 1 and 2). A person whose logical stage is only partially formal operational is limited to the conventional moral stages (Stages 3 and 4). While logical development is necessary for moral development and sets limits to it, most individuals are higher in logical stage than they are in moral stage. As an example, over 50% of late adolescents and adults are capable of full formal reasoning, but only 10% of these adults (all formal operational) display principled (Stages 5 and 6) moral reasoning.

The moral stages are *structures of moral judgment* or *moral reasoning. Structures* of moral judgment must be distinguished from the *content* of moral judgment. As an example, we cite responses to a dilemma used in our various studies to identify moral stage. The dilemma raises the issue of stealing a drug to save a dying woman. The inventor of the drug is selling it for 10 times what it costs him to make it. The woman's husband cannot raise the money, and the seller refuses to lower the price or wait for payment. What should the husband do?

The choice endorsed by a subject (steal, don't steal) is called the *content* of his moral judgment in the situation. His reasoning about the choice defines the structure of his moral judgment. This reasoning centers on the following 10 universal moral values or issues of concern to persons in these moral dilemmas:

1. Punishment
2. Property
3. Roles and concerns of affection
4. Roles and concerns of authority
5. Law
6. Life
7. Liberty
8. Distributive justice
9. Truth
10. Sex

A moral choice involves choosing between two (or more) of these values as they *conflict* in concrete situations of choice.

The stage or structure of a person's moral judgment defines: 1) *what* he finds valuable in each of these moral issues (life, law), i.e., how he defines the value, and 2) *why* he finds it valuable, i.e., the reasons he gives for valuing it. As an example, at Stage 1 life is valued in terms of the power or possessions of the person involved; at Stage 2, for its usefulness in satisfying the needs of the individual in question or others; at Stage 3, in terms of the individual's relations with others and their valuation of him; at Stage 4, in

terms of social or religious law. Only at Stages 5 and 6 is each life seen as inherently worthwhile, aside from other considerations.

Moral judgment vs. moral action

Having clarified the nature of stages of moral *judgment*, we must consider the relation of moral judgment to moral *action*. If logical reasoning is a necessary but not sufficient condition for mature moral judgment, mature moral judgment is a necessary but not sufficient condition for mature moral action. One cannot follow moral principles if one does not understand (or believe in) moral principles. However, one can reason in terms of principles and not live up to these principles. As an example, Richard Krebs and I found that only 15% of students showing some principled thinking cheated as compared to 55% of conventional subjects and 70% of preconventional subjects.[5] Nevertheless, 15% of the principled subjects did cheat, suggesting that factors additional to moral judgment are necessary for principled moral reasoning to be translated into "moral action." Partly, these factors include the situation and its pressures. Partly, what happens depends upon the individual's motives and emotions. Partly, what the individual does depends upon a general sense of will, purpose, or "ego strength." As an example of the role of will or ego strength in moral behavior, we may cite the study by Krebs: Slightly more than half of his conventional subjects cheated. These subjects were also divided by a measure of attention/will. Only 26% of the "strong-willed" conventional subjects cheated; however, 74% of the "weak-willed" subjects cheated.

If maturity of moral reasoning is only one factor in moral behavior, why does the cognitive-developmental approach to moral education focus so heavily upon moral reasoning? For the following reasons:

1. Moral judgment, while only one factor in moral behavior, is the single most important or influential factor yet discovered in moral behavior.
2. While other factors influence moral behavior, moral judgment is the only distinctively *moral* factor in moral behavior. To illustrate, we noted that the Krebs study indicated that "strong-willed" conventional stage subjects resisted cheating more than "weak-willed" subjects. For those at a preconventional level of moral reasoning, however, "will" had an opposite effect. "Strong-willed" Stages 1 and 2 subjects cheated more, not less, than "weak-willed" subjects, i.e., they had the "courage of their (amoral) convictions" that it was worthwhile to cheat. "Will," then, is an important factor in moral behavior, but it is not distinctively moral; it becomes moral only when informed by mature moral judgment.
3. Moral judgment change is long-range or irreversible; a higher stage is never lost. Moral behavior as such is largely situational and reversible or "loseable" in new situations.

II. Aims of moral and civic education

Moral psychology describes what moral development is, as studied empirically. Moral education must also consider moral philosophy, which strives to tell us what moral development ideally *ought to be*. Psychology finds an invariant sequence of moral stages; moral philosophy must be invoked to answer whether a later stage is a better stage. The "stage" of senescence and death follows the "stage" of adulthood, but that does not mean that senescence and death are better. Our claim that the latest or principled stages of moral reasoning are morally better stages, then, must rest on considerations of moral philosophy.

The tradition of moral philosophy to which we appeal is the liberal or rational tradition, in particular the "formalistic" or "deontological" tradition running from Immanuel Kant to John Rawls.[6] Central to this tradition is the claim that an adequate morality is *principled*, i.e., that it makes judgments in terms of *universal* principles applicable to all mankind. *Principles* are to be distinguished from *rules*. Conventional morality is grounded on rules, primarily "thou shalt nots" such as are represented by the Ten Commandments, prescriptions of kinds of actions. Principles are, rather, universal guides to making a moral decision. An example is Kant's "categorical imperative," formulated in two ways. The first is the maxim of respect for human personality, "Act always toward the other as an end, not as a means." The second is the maxim of universalization, "Choose only as you would be willing to have everyone choose in your situation." Principles like that of Kant's state the formal conditions of a moral choice or action. In the dilemma in which a woman is dying because a druggist refuses to release his drug for less than the stated price, the druggist is not acting morally, though he is not violating the ordinary moral rules (he is not actually stealing or murdering). But he is violating principles: He is treating the woman simply as a means to his ends of profit, and he is not choosing as he would wish anyone to choose (if the druggist were in the dying woman's place, he would not want a druggist to choose as he is choosing). Under most circumstances, choice in terms of conventional moral rules and choice in terms of principles coincide. Ordinarily, principles dictate not stealing (avoiding stealing is implied by acting in terms of a regard for others as ends and in terms of what one would want everyone to do). In a situation where stealing is the only means to save a life, however, principles contradict the ordinary rules and would dictate stealing. Unlike rules which are supported by social authority, principles are freely chosen by the individual because of their intrinsic moral validity.[d]

The conception that a moral choice is a choice made in terms of moral principles is related to the claim of liberal moral philosophy that moral principles are ultimately principles of justice. In essence, moral conflicts are conflicts between the claims of persons, and principles for resolving these

claims are principles of justice, "for giving each his due." Central to justice are the demands of *liberty, equality*, and *reciprocity*. At every moral stage, there is a concern for justice. The most damning statement a school child can make about a teacher is that "he's not fair." At each higher stage, however, the conception of justice is reorganized. At Stage 1, justice is punishing the bad in terms of "an eye for an eye and a tooth for a tooth." At Stage 2, it is exchanging favors and goods in an equal manner. At Stages 3 and 4, it is treating people as they desire in terms of the conventional rules. At Stage 5, it is recognized that all rules and laws flow from justice, from a social contract between the governors and the governed designed to protect the equal rights of all. At Stage 6, personally chosen moral principles are also principles of justice, the principles any member of a society would choose for that society if he did not know what his position was to be in the society and in which he might be the least advantaged.[7] Principles chosen from this point of view are, first, the maximum liberty compatible with the like liberty of others and, second, no inequalities of goods and respect which are not to the benefit of all, including the least advantaged.

As an example of stage progression in the orientation to justice, we may take judgments about capital punishment.[8] Capital punishment is only firmly rejected at the two principled stages, when the notion of justice as vengeance or retribution is abandoned. At the sixth stage, capital punishment is not condoned even if it may have some useful deterrent effect in promoting law and order. This is because it is not a punishment we would choose for a society if we assumed we had as much chance of being born into the position of a criminal or murderer as being born into the position of a law abider.

Why are decisions based on universal principles of justice better decisions? Because they are decisions on which all moral men could agree. When decisions are based on conventional moral rules, men will disagree, since they adhere to conflicting systems of rules dependent on culture and social position. Throughout history men have killed one another in the name of conflicting moral rules and values, most recently in Vietnam and the Middle East. Truly moral or just resolutions of conflicts require principles which are, or can be, universalizable.

Alternative approaches

We have given a philosophic rationale for stage advance as the aim of moral education. Given this rationale, the developmental approach to moral education can avoid the problems inherent in the other two major approaches to moral education. The first alternative approach is that of indoctrinative moral education, the preaching and imposition of the rules and values of the teacher and his culture on the child. In America, when this indoctrinative approach has been developed in a systematic manner, it has usually been termed "character education."

Moral values, in the character education approach, are preached or taught in terms of what may be called the "bag of virtues." In the classic studies of character by Hugh Hartshorne and Mark May, the virtues chosen were honesty, service, and self-control.[9] It is easy to get superficial consensus on such a bag of virtues – until one examines in detail the list of virtues involved and the details of their definition. Is the Hartshorne and May bag more adequate than the Boy Scout bag (a Scout should be honest, loyal, reverent, clean, brave, etc.)? When one turns to the details of defining each virtue, one finds equal uncertainty or difficulty in reaching consensus. Does honesty mean one should not steal to save a life? Does it mean that a student should not help another student with his homework?

Character education and other forms of indoctrinative moral education have aimed at teaching universal values (it is assumed that honesty or service are desirable traits for all men in all societies), but the detailed definitions used are relative; they are defined by the opinions of the teacher and the conventional culture and rest on the authority of the teacher for their justification. In this sense character education is close to the unreflective valuings by teachers which constitute the hidden curriculum of the school.[e] Because of the current unpopularity of indoctrinative approaches to moral education, a family of approaches called "values clarification" has become appealing to teachers. Values clarification takes the first step implied by a rational approach to moral education: the eliciting of the child's own judgment or opinion about issues or situations in which values conflict, rather than imposing the teacher's opinion on him. Values clarification, however, does not attempt to go further than eliciting awareness of values; it is assumed that becoming more self-aware about one's values is an end in itself. Fundamentally, the definition of the end of values education as self-awareness derives from a belief in ethical relativity held by many value-clarifiers. As stated by Peter Engel, "One must contrast value clarification and value inculcation. Value clarification implies the principle that in the consideration of values there is no single correct answer." Within these premises of "no correct answer," children are to discuss moral dilemmas in such a way as to reveal different values and discuss their value differences with each other. The teacher is to stress that "our values are different," not that one value is more adequate than others. If this program is systematically followed, students will themselves become relativists, believing there is no "right" moral answer. For instance, a student caught cheating might argue that he did nothing wrong, since his own hierarchy of values, which may be different from that of the teacher, made it right for him to cheat.

Like values clarification, the cognitive-developmental approach to moral education stresses open or Socratic peer discussion of value dilemmas. Such discussion, however, has an aim: stimulation of movement to the next stage of moral reasoning. Like values clarification, the developmental approach opposes indoctrination. Stimulation of movement to the next stage of reasoning is not indoctrinative, for the following reasons:

1. Change is in the way of reasoning rather than in the particular beliefs involved.
2. Students in a class are at different stages; the aim is to aid movement of each to the next stage, not convergence on a common pattern.
3. The teacher's own opinion is neither stressed nor invoked as authoritative. It enters in only as one of many opinions, hopefully one of those at a next higher stage.
4. The notion that some judgments are more adequate than others is communicated. Fundamentally, however, this means that the student is encouraged to articulate a position which seems most adequate to him and to judge the adequacy of the reasoning of others.

In addition to having more definite aims than values clarification, the moral development approach restricts value education to that which is moral or, more specifically, to justice. This is for two reasons. First, it is not clear that the whole realm of personal, political, and religious values is a realm which is nonrelative, i.e., in which there are universals and a direction of development. Second, it is not clear that the public school has a right or mandate to develop values in general.[f] In our view, value education in the public schools should be restricted to that which the school has the right and mandate to develop: an awareness of justice, or of the rights of others in our Constitutional system. While the Bill of Rights prohibits the teaching of religious beliefs, or of specific value systems, it does not prohibit the teaching of the awareness of rights and principles of justice fundamental to the Constitution itself.

When moral education is recognized as centered in justice and differentiated from value education or affective education, it becomes apparent that moral and civic education are much the same thing. This equation, taken for granted by the classic philosophers of education from Plato and Aristotle to Dewey, is basic to our claim that a concern for moral education is central to the educational objectives of social studies.

The term *civic education* is used to refer to social studies as more than the study of the facts and concepts of social science, history, and civics. It is education for the analytic understanding, value principles, and motivation necessary for a citizen in a democracy if democracy is to be an effective process. It is political education. Civic or political education means the stimulation of development of more advanced patterns of reasoning about political and social decisions and their implementation in action. These patterns are patterns of moral reasoning. Our studies show that reasoning and decision making about political decisions are directly derivative of broader patterns of moral reasoning and decision making. We have interviewed high school and college students about concrete political situations involving laws to govern open housing, civil disobedience for peace in Vietnam, free press rights to publish what might disturb national order, and distribution

of income through taxation. We find that reasoning on these political decisions can be classified according to moral stage and that an individual's stage on political dilemmas is at the same level as on nonpolitical moral dilemmas (euthanasia, violating authority to maintain trust in a family, stealing a drug to save one's dying wife). Turning from reasoning to action, similar findings are obtained. In 1963 a study was made of those who sat in at the University of California, Berkeley, administration building and those who did not in the Free Speech Movement crisis. Of those at Stage 6, 80% sat in, believing that principles of free speech were being compromised, and that all efforts to compromise and negotiate with the administration had failed. In contrast, only 15% of the conventional (Stage 3 or Stage 4) subjects sat in. (Stage 5 subjects were in between.)[g]

From a psychological side, then, political development is part of moral development. The same is true from the philosophic side. In the *Republic*, Plato sees political education as part of a broader education for moral justice and finds a rationale for such education in terms of universal philosophic principles rather than the demands of a particular society. More recently, Dewey claims the same.

In historical perspective, America was the first nation whose government was publicly founded on postconventional principles of justice, rather than upon the authority central to conventional moral reasoning. At the time of our founding, postconventional or principled moral and political reasoning was the possession of the minority, as it still is. Today, as in the time of our founding, the majority of our adults are at the conventional level, particualrly the "law and order" (fourth) moral stage. (Every few years the Gallup Poll circulates the Bill of Rights unidentified, and every year it is turned down.) The Founding Fathers intuitively understood this without benefit of our elaborate social science research; they constructed a document designing a government which would maintain principles of justice and the rights of man even though principled men were not the men in power. The machinery included checks and balances, the independent judiciary, and freedom of the press. Most recently, this machinery found its use at Watergate. The tragedy of Richard Nixon, as Harry Truman said long ago, was that he never understood the Constitution (a Stage 5 document), but the Constitution understood Richard Nixon.[h]

Watergate, then, is not some sign of moral decay of the nation, but rather of the fact that understanding and action in support of justice principles are still the possession of a minority of our society. Insofar as there is moral decay, it represents the weakening of conventional morality in the face of social and value conflict today. This can lead the less fortunate adolescent to fixation at the preconventional level, the more fortunate to movement to principles. We find a larger proportion of youths at the principled level today than was the case in their fathers' day, but also a larger proportion at the preconventional level.

Given this state, moral and civic education in the schools becomes a more urgent task. In the high school today, one often hears both preconventional adolescents and those beginning to move beyond convention sounding the same note of disaffection for the school. While our political institutions are in principle Stage 5 (i.e., vehicles for maintaining universal rights through the democratic process), our schools have traditionally been Stage 4 institutions of convention and authority. Today more than ever, democratic schools systematically engaged in civic education are required.

Our approach to moral and civic education relates the study of law and government to the actual creation of a democratic school in which moral dilemmas are discussed and resolved in a manner which will stimulate moral development.

Planned moral education

For many years, moral development was held by psychologists to be primarily a result of family upbringing and family conditions. In particular, conditions of affection and authority in the home were believed to be critical, some balance of warmth and firmness being optimal for moral development. This view arises if morality is conceived as an internalization of the arbitrary rules of parents and culture, since such acceptance must be based on affection and respect for parents as authorities rather than on the rational nature of the rules involved.

Studies of family correlates of moral stage development do not support this internalization view of the conditions for moral development. Instead, they suggest that the conditions for moral development in homes and schools are similar and that the conditions are consistent with cognitive-developmental theory. In the cognitive-developmental view, morality is a natural product of a universal human tendency toward empathy or role taking, toward putting oneself in the shoes of other conscious beings. It is also a product of a universal human concern for justice, for reciprocity or equality in the relation of one person to another. As an example, when my son was 4, he became a morally principled vegetarian and refused to eat meat, resisting all parental persuasion to increase his protein intake. His reason was, "It's bad to kill animals." His moral commitment to vegetarianism was not taught or acquired from parental authority; it was the result of the universal tendency of the young self to project its consciousness and values into other living things, other selves. My son's vegetarianism also involved a sense of justice, revealed when I read him a book about Eskimos in which a real hunting expedition was described. His response was to say, "Daddy, there is one kind of meat I would eat – Eskimo meat. It's all right to eat Eskimos because they eat animals." This natural sense of justice or reciprocity was Stage 1 – an eye for an eye, a tooth for a tooth. My son's sense of the value of life was also Stage 1 and involved no differentiation between

human personality and physical life. His morality, though Stage 1, was, however, natural and internal. Moral development past Stage 1, then, is not an internalization but the reconstruction of role taking and conceptions of justice toward greater adequacy. These reconstructions occur in order to achieve a better match between the child's own moral structures and the structures of the social and moral situations he confronts. We divide these conditions of match into two kinds: those dealing with moral discussions and communication and those dealing with the total moral environment or atmosphere in which the child lives.

In terms of moral discussion, the important conditions appear to be:

1. Exposure to the next higher stage of reasoning
2. Exposure to situations posing problems and contradictions for the child's current moral structure, leading to dissatisfaction with his current level
3. An atmosphere of interchange and dialogue combining the first two conditions, in which conflicting moral views are compared in an open manner

Studies of families in India and America suggest that morally advanced children have parents at higher stages. Parents expose children to the next higher stage, raising moral issues and engaging in open dialogue or interchange about such issues.[10]

Drawing on this notion of the discussion conditions stimulating advance, Moshe Blatt conducted classroom discussions of conflict-laden hypothetical moral dilemmas with four classes of junior high and high school students for a semester.[11] In each of these classes, students were to be found at three stages. Since the children were not all responding at the same stage, the arguments they used with each other were at different levels. In the course of these discussions among the students, the teacher first supported and clarified those arguments that were one stage above the lowest stage among the children; for example, the teacher supported Stage 3 rather than Stage 2. When it seemed that these arguments were understood by the students, the teacher then challenged that stage, using new situations, and clarified the arguments one stage above the previous one: Stage 4 rather than Stage 3. At the end of the semester, all the students were retested; they showed significant upward change when compared to the controls, and they maintained the change one year later. In the experimental classrooms, from one-fourth to one-half of the students moved up a stage, while there was essentially no change during the course of the experiment in the control group.

Given the Blatt studies showing that moral discussion could raise moral stage, we undertook the next step: to see if teachers could conduct moral discussions in the course of teaching high school social studies with the same results. This step we took in cooperation with Edwin Fenton, who introduced moral dilemmas in his ninth- and eleventh-grade social studies texts. Twenty-four teachers in the Boston and Pittsburgh areas were given some

instruction in conducting moral discussions around the dilemmas in the text. About half of the teachers stimulated significant developmental change in their classrooms – upward stage movement of one-quarter to one-half a stage. In control classes using the text but no moral dilemma discussions, the same teachers failed to stimulate any moral change in the students. Moral discussion, then, can be a usable and effective part of the curriculum at any grade level. Working with filmstrip dilemmas produced in cooperation with Guidance Associates, second-grade teachers conducted moral discussions yielding a similar amount of moral stage movement.

Moral discussion and curriculum, however, constitute only one portion of the conditions stimulating moral growth. When we turn to analyzing the broader life environment, we turn to a consideration of the *moral atmosphere* of the home, the school, and the broader society. The first basic dimension of social atmosphere is the role-taking opportunities it provides, the extent to which it encourages the child to take the point of view of others. Role taking is related to the amount of social interaction and social communication in which the child engages, as well as to his sense of efficacy in influencing attitudes of others. The second dimension of social atmosphere, more strictly moral, is the level of justice of the environment or institution. The justice structure of an institution refers to the perceived rules or principles for distributing rewards, punishments, responsibilities, and privileges among institutional members. This structure may exist or be perceived at any of our moral stages. As an example, a study of a traditional prison revealed that inmates perceived it as Stage 1, regardless of their own level.[12] Obedience to arbitrary command by power figures and punishment for disobedience were seen as the governing justice norms of the prison. A behavior-modification prison using point rewards for conformity was perceived as a Stage 2 system of instrumental exchange. Inmates at Stage 3 or 4 perceived this institution as more fair than the traditional prison, but not as fair in their own terms.

These and other studies suggest that a higher level of institutional justice is a condition for individual development of a higher sense of justice. Working on these premises, Joseph Hickey, Peter Scharf, and I worked with guards and inmates in a women's prison to create a more just community.[13] A social contract was set up in which guards and inmates each had a vote of one and in which rules were made and conflicts resolved through discussions of fairness and a democratic vote in a community meeting. The program has been operating four years and has stimulated moral stage advance in inmates, though it is still too early to draw conclusions as to its overall long-range effectiveness for rehabilitation.

One year ago, Fenton, Ralph Mosher, and I received a grant from the Danforth Foundation (with additional support from the Kennedy Foundation) to make moral education a living matter in two high schools in the Boston area (Cambridge and Brookline) and two in Pittsburgh. The plan

had two components. The first was training counselors and social studies and English teachers in conducting moral discussions and making moral discussion an integral part of the curriculum. The second was establishing a just community school within a public high school.

We have stated the theory of the just community high school, postulating that discussing real-life moral situations and actions as issues of fairness and as matters for democratic decision would stimulate advance in both moral reasoning and moral action. A participatory democracy provides more extensive opportunities for role taking and a higher level of perceived institutional justice than does any other social arrangement. Most alternative schools strive to establish a democratic governance, but none we have observed has achieved a vital or viable participatory democracy. Our theory suggested reasons why we might succeed where others failed. First, we felt that democracy had to be a central commitment of a school, rather than a humanitarian frill. Democracy as moral education provides that commitment. Second, democracy in alternative schools often fails because it bores the students. Students prefer to let teachers make decisions about staff, courses, and schedules, rather than to attend lengthy, complicated meetings. Our theory said that the issues a democracy should focus on are issues of morality and fairness. Real issues concerning drugs, stealing, disruptions, and grading are never boring if handled as issues of fairness. Third, our theory told us that if large democratic community meetings were preceded by small-group moral discussion, higher-stage thinking by students would win out in later decisions, avoiding the disasters of mob rule.[i]

Currently, we can report that the school based on our theory makes democracy work or function where other schools have failed. It is too early to make any claims for its effectiveness in causing moral development, however.

Our Cambridge just community school within the public high school was started after a small summer planning session of volunteer teachers, students, and parents. At the time the school opened in the fall, only a commitment to democracy and a skeleton program of English and social studies had been decided on. The school started with six teachers from the regular school and 60 students, 20 from academic professional homes and 20 from working-class homes. The other 20 were dropouts and troublemakers or petty delinquents in terms of previous record. The usual mistakes and usual chaos of a beginning alternative school ensued. Within a few weeks, however, a successful democratic community process had been established. Rules were made around pressing issues: disturbances, drugs, hooking. A student discipline committee or jury was formed. The resulting rules and enforcement have been relatively effective and reasonable. We do not see reasonable rules as ends in themselves, however, but as vehicles for moral discussion and an emerging sense of community. This sense of community and a resulting morale are perhaps the most immediate signs of success. This sense of community seems to lead to behavior change of a positive sort. An example

is a 15-year-old student who started as one of the greatest combinations of humor, aggression, light-fingeredness, and hyperactivity I have ever known. From being the principal disturber of all community meetings, he has become an excellent community meeting participant and occasional chairman. He is still more ready to enforce rules for others than to observe them himself, yet his commitment to the school has led to a steady decrease in exotic behavior. In addition, he has become more involved in classes and projects and has begun to listen and ask questions in order to pursue a line of interest.

We attribute such behavior change not only to peer pressure and moral discussion but to the sense of community which has emerged from the democratic process in which angry conflicts are resolved through fairness and community decision. This sense of community is reflected in statements of the students to us that there are no cliques – that the blacks and the whites, the professors' sons and the project students, are friends. These statements are supported by observation. Such a sense of community is needed where students in a given classroom range in reading level from fifth-grade to college.

Fenton, Mosher, the Cambridge and Brookline teachers, and I are now planning a four-year curriculum in English and social studies centering on moral discussion, on role taking and communication, and on relating the government, laws, and justice system of the school to that of the American society and other world societies. This will integrate an intellectual curriculum for a higher level of understanding of society with the experiential components of school democracy and moral decision.

There is very little new in this – or in anything else we are doing. Dewey wanted democratic experimental schools for moral and intellectual development 70 years ago. Perhaps Dewey's time has come.

Notes

1 John Dewey, "What Psychology Can Do for the Teacher," in Reginald Archambault, ed., *John Dewey on Education: Selected Writings* (New York: Random House, 1964).
2 Jean Piaget, *The Moral Judgment of the Child*, 2nd ed. (Glencoe, Ill.: Free Press, 1948).
3 Lawrence Kohlberg, "Moral Stages and Moralization: The Cognitive-Developmental Approach," in Thomas Lickona, ed., *Man, Morality, and Society* (New York: Holt, Rinehart, and Winston, in press).
4 James Rest, Elliott Turiel, and Lawrence Kohlberg, "Relations Between Level of Moral Judgment and Preference and Comprehension of the Moral Judgment of Others," *Journal of Personality*, vol. 37, 1969, pp. 225–52, and James Rest, "Comprehension, Preference, and Spontaneous Usage in Moral Judgment," in Lawrence Kohlberg, ed., *Recent Research in Moral Development* (New York: Holt, Rinehart, and Winston, in preparation).
5 Richard Krebs and Lawrence Kohlberg, "Moral Judgment and Ego Controls as Determinants of Resistance to Cheating," in Lawrence Kohlberg, ed., *Recent Research*.

6 John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971).
7 John Rawls, ibid.
8 Lawrence Kohlberg and Donald Elfenbein, "Development of Moral Reasoning and Attitudes Toward Capital Punishment," *American Journal of Orthopsychiatry*, Summer, 1975.
9 Hugh Hartshorne and Mark May, *Studies in the Nature of Character: Studies in Deceit*, vol. 1; *Studies in Service and Self-Control*, vol. 2; *Studies in Organization of Character*, vol. 3 (New York: Macmillan, 1928–30).
10 Bindu Parilch, "A Cross-Cultural Study of Parent-Child Moral Judgment," unpublished doctoral dissertation, Harvard University, 1975.
11 Moshe Blatt and Lawrence Kohlberg, "Effects of Classroom Discussions upon Children's Level of Moral Judgment," in Lawrence Kohlberg, ed., *Recent Research*.
12 Lawrence Kohlberg, Peter Scharf, and Joseph Hickey, "The Justice Structure of the Prison: A Theory and an Intervention," *The Prison Journal*, Autumn-Winter, 1972.
13 Lawrence Kohlberg, Kelsey Kauffman, Peter Scharf, and Joseph Hickey, *The Just Community Approach to Corrections: A Manual, Part I* (Cambridge, Mass.: Education Research Foundation, 1973).

Notes

a These levels correspond roughly to our three major levels: the preconventional, the conventional, and the principled. Similar levels were propounded by William McDougall, Leonard Hobhouse, and James Mark Baldwin.

b Piaget's stages correspond to our first three stages: Stage 0 (pre-moral), Stage 1 (heteronomous), and Stage 2 (instrumental reciprocity).

c Many adolescents and adults only partially attain the stage of formal operations. They do consider all the actual relations of one thing to another at the same time, but they do not consider all possibilities and form abstract hypotheses. A few do not advance this far, remaining "concrete operational."

d Not all freely chosen values or rules are principles, however. Hitler chose the "rule," "exterminate the enemies of the Aryan race," but such a rule is not a universalizable principle.

e As an example of the "hidden curriculum," we may cite a second-grade classroom. My son came home from this classroom one day saying he did not want to be "one of the bad boys." Asked "Who are the bad boys?" he replied, "The ones who don't put their books back and get yelled at."

f Restriction of deliberate value education to the moral may be clarified by our example of the second-grade teacher who made tidying up of books a matter of moral indoctrination. Tidiness is a value, but it is not a moral value. Cheating is a moral issue, intrinsically one of fairness. It involves issues of violation of trust and taking advantage. Failing to tidy the room may under certain conditions be an issue of fairness, when it puts an undue burden on others. If it is handled by the teacher as a matter of cooperation among the group in this sense, it is a legitimate focus of deliberate moral education. If it is not, it simply represents the arbitrary imposition of the teacher's values on the child.

g The differential action of the principled subjects was determined by two things. First, they were more likely to judge it right to violate authority by sitting in. But second, they were also in general more consistent in engaging in political action according to their judgment. Ninety percent of all Stage 6 subjects thought it right to sit in, and all 90% lived up to this belief. Among the Stage 4 subjects, 45% thought it right to sit in, but only 33% lived up to this belief by acting.

h No public or private word or deed of Nixon ever rose above Stage 4, the "law and order" stage. His last comments in the White House were of wonderment that the Republican Congress could turn on him after so many Stage 2 exchanges of favors in getting them elected.

i An example of the need for small-group discussion comes from an alternative school community meeting called because a pair of the students had stolen the school's video-recorder. The resulting majority decision was that the school should buy back the recorder from the culprits through a fence. The teachers could not accept this decision and returned to a more authoritative approach. I believe if the moral reasoning of students urging this solution had been confronted by students at a higher stage, a different decision would have emerged.

89

KOHLBERG'S DORMANT GHOSTS

The case of education

F. K. Oser

Moral developmental theory: an unchallenged cornerstone

In one of the Swiss cantons, people tell a legend about a ghost who always helps the farmers if they get into any kind of trouble. The ghost is omnipresent: he keeps an eye on bad behaviour, helps the poor, and exerts a hidden power that informs every human endeavour. So much for the positive side of the ghost. The negative side is that this ghost prevents people from being autonomous. He is omnipotent, and his fundamental strength is so great that humans do not ever seem to be able to liberate themselves from his power.

There is something about Kohlberg's legacy which is similar to the ambivalent ghost just mentioned. On one hand, Kohlberg's theory is extremely helpful for thinking in terms of moral development. On the other hand, the theory of moral stages provides such a strong and powerful paradigm that the theory or its major assumptions seem almost impossible to falsify. To give an example: one of our German colleagues, an outstanding scholar, tried to show that at least the so-called segmentation gap with respect to subjects such as apprentices and professionals is bigger than Kohlberg assumed. For the time being, however, his research data do not support him in his opinion (Beck, in preparation).

Thus, no one doubts that Lawrence Kohlberg's basic theory is an unchallenged and today widely accepted cornerstone in this research field, not only with respect to the stage theory but also concerning its educational implications and applications such as the dilemma discussion approach to classroom moral education, curriculum development, the Just Community approach in different settings, and developmental counselling. Even if we do accept new and different measurement methods (Eckensberger & Reinshagen, 1978; Lind, 1984; Rest & Narvez, 1994) that provided some differentiation of the theory, and even if we do see that attempts to reconstruct the stage hierarchy lead to revised definitions of Stage One and Two for instance (see Keller, 1990;

Source: *Journal of Moral Education*, 1996, 25(3), 253–275.

Teo *et al.*, 1995), most of us will agree that Kohlberg's theory was an enormous step beyond that of Piaget, that it was well conceived with regard to its philosophical foundation, and that it was empirically well tested. Moreover, many of the critics did not really see the degree of elaboration of the theory provided by Kohlberg's followers. This holds true, for example, for the field of social development, the development of the concept of friendship, the development of the concept of authority, and religious development (see Damon, 1977; Selman, 1980; Fowler, 1981; Keller, 1984; Edelstein & Keller, 1986; Oser & Gmünder, 1991; Oser & Scarlett, 1991, etc.). In addition, domain specific development (Nucci, 1981; Turiel, 1983; Nisan, 1987) and research on the gap of judgement and action (Blasi, 1980, 1983) provided valid answers to important and critical questions. In this research tradition that started with Kohlberg, hundreds of studies have been conducted; in addition, we have experienced an ongoing growth of educational concern to which the Kohlberg approach, more than any other psychological paradigm, is still a convincing answer.

Unanswered questions

However, I am talking about *dormant* ghosts; that means that within the frame of Kohlberg's theory many questions have been left unanswered. Kohlberg himself already saw, or rather *felt*, many unsolved problems. Reading Kohlberg closely, we have to realise that his thinking intuitively brought up much more than his writing finally presents: black holes of his work that I would like to nickname the "dormant ghosts". In what follows, I would like to wake up some of them a little because in many ways they can help us to start new and convincing research programmes. I will especially focus on moral education rather than on moral development itself. This is why I leave out many classical unsolved questions such as: the judgement action gap, measurement problems in general, domain specificity, the ontological status of Stage Seven, intercultural comparisons, the separation of orientations of justice and care, the reconstruction of the Stage One and Two, the influence of new information on decision making, and so forth.

The Just Community approach

One of Kohlberg's great educational achievements is the creation of the Just Community school reform. For the field of moral education this concept is similar to a "quantum leap". No one since Janusz Korczak has engaged himself with so much effort to make a whole school a matter of developmentally orientated moral education, using "round table" situations in order to solve concrete social and moral problems and, thus, learning morally. Kohlberg's idea of schools becoming small democracies is a moral endeavour that relies on the notion that to allocate responsibility means to create,

somehow "officially", a link between moral judgement and moral action; this concept reintroduced meaning-making as the major target of developmental school reform. All psychological effects—such as the disequilibrium and stimulation of cognitive structures, role-taking pressure, interruption of school flow in order to listen to claims and needs of students, identification with rules through common rule creation and decision-making, taking responsibility and caring for fellow students—all these important and fundamental elements of moral education are holistically united in this vision of a Just Community. In addition, the institution as a whole cares to have a positive value. The notion of collectiveness (Power *et al*., 1989) and the notion of moral climate (Lind & Link, 1986) are basic elements of a growing awareness of the child and adolescent that the school can become a major issue in their life and the life of all the others. Stages, phases and types of norms relate to different forms of action and different connections between judgement and action (see Power *et al*., 1989).

Included in the Just Community process were all the experiences with attempts and errors *before* its creation, that is, the studies on the stimulation of a higher stage (from Blatt & Kohlberg, 1975, until Rest & Narvaez, 1994). Indeed, this was a very painful process. There were so many unsolved problems, such as the argument on the optimal conditions of the proximal zone of development (+ 1 condition, + $^1/_2$ condition: see Berkowitz *et al*., 1980); the question of cognitive moral development with or without emotional stimulation; and the question of the curriculum, which remained completely unsolved (positive exceptions being *Man. A Course of Study* by Bruner, 1966, or *Facing History and Ourselves*, a curriculum from Stern-Strom, 1980). There were also many good ideas that led to new problems: for instance, the belief that content does not play any important role in the stimulation of a higher moral structure has led to an almost complete neglect of content matters, to what we call a decontextualisation of morality. In addition, one of the most serious problems was that teachers did not know how to integrate the technique of dilemma discussions into the "normal" instruction strategies within the different subjects. They know how to transmit knowledge but not how to stimulate a moral disequilibrium. Perhaps the most problematic issue was that students did not relate moral judgement to action, and neither shame nor indignation were coupled to the discourse of rightness within the dilemma discussion. Not to keep a promise was a question of debate but not a question of caring for people or for fine-tuned interpersonal relationships, with their necessary acts of apologising and reparation.

When in 1974/75 the idea of the Just Community was generated in collaboration with teachers, principals and administrators, Kohlberg perceived schools as places where all these problems could be solved. Indeed, by this Dewey-orientated social form of life "in which interests are mutually interpenetrating and where progress, or readjustments systems, is an important consideration" (Dewey, 1944, p. 87), many of these weaknesses have been

overcome. Dewey himself talked about starting with a narrowly perceived morality which relies on "a sentimental goody-goody turn without reference to effective ability to do what is socially needed" (1944, p. 357) and moving towards a democratic school setting; and Kohlberg left his former focus on the mere stimulation of higher stage reasoning behind and developed a communitarian form of moral education.

It is interesting that Kohlberg, at that time about 50 years of age, entered this new pedagogical field with a total commitment. Before this practical work he did not have to soil his hands—he was a pure cognitive–developmental psychologist. However, Kohlberg soiled his hands when he went out to poor schools, to prisons and to difficult urban settings in order to change moral feelings, acts and judgements in schools as a whole, and also the individuals who were supposed to transform the school into a community. Here he encountered action problems, measurement problems, political problems and many insecurities.

Why is it so important that schools become communities? In discussing and arguing about issues of school life in the public sphere of the school, students learn about the importance of the public sphere of discourse. They learn that what seems to concern just one or two students—for instance, when a bike was stolen—in reality should be a concern of all the students of a school—a learning process that represents a first form of universalisation. They learn what it means to justify behaviours in front of classmates and all the other students. These aspects are a first and very important step towards role taking and universal moral thinking. Secondly, students who are allowed to participate in decision making about the social rules of their school develop an identification and greater commitment with respect to these rules, since they helped to establish them. Such rules are, at least in part, rules and school norms of the student concerned. School is then less likely to be regarded as some sort of academic shopping mall where everybody simply looks for his or her own profits and benefits—it becomes, as several studies were able to show, more likely to be seen as "my school" and "our school"; it becomes an object of pride.

Schooling is, then, an holistic endeavour. It is difficult to disentangle the setting, and, viewed from a sociological perspective, the community can become morally better/higher than the average morality of the individuals within it. The whole is more than the sum of its parts.

A European version of the Just Community approach

In 1985 the minister of education of the German state of North Rhine Westfalia made an open statement for moral education in German schools. This statement was perceived as ambiguous by many teachers, unions and parents because it seemed to call for a traditional value transmission curriculum. An initiative by Georg Lind led the minister to invite Lawrence

Kohlberg and Ann Higgins to come to Germany for a presentation of their model. Shortly after, the way was open for installing Just Communities in some schools in North Rhine Westfalia. In 1987, an Advisory Board was established with deputies from the government, teachers, inspectors and scientific consultants such as G. Lind, H. Schirp and the author. After long discussions on how to apply the models, Lind, Schirp and I presented our suggestions to different schools (including information with respect to Kohlberg's developmental theory). Three of the schools decided to participate. In 1987, the project was put into practice.

What was the general frame of the project? Like the American model, the German Just Community is an approach that gives students and teachers the possibility of regulating the inner life of a school themselves through a democratic decision making process and thus to learn morally and socially. The Just Community methods generate conditions for a comprehensive moral culture of a school, based on a lived appeal to the reasoning capacity of every member of the community (see Power *et al*., 1989, p. 33 ff).

The Just Community approach puts forward a few fundamental goals which are connected to one another holistically. These goals are:

- the creation and adaptation of justified shared rules by all participants (solidarity); stimulation of moral judgement competence;
- maintenance of the congruence between moral judgement and moral action;
- training of moral empathy and encouragement of prosocial commitment; and
- development of a solid value system based on tolerance and openness (see Lempert, 1988; Power *et al*., 1989; Oser & Althof, 1992).

In contrast to the American model, the German version insisted on a number of additional considerations:

1. Teachers and principals have to be thoroughly prepared for a transformation in their conception of authority. In-service training always took place after the Just Community meetings (which define the core of the model) and was designed to evaluate the role of the faculty and the individual teacher during these meetings.
2. The second new dimension of the German model was that the children were significantly younger than in the American programmes; we started working with 5th graders, i.e. 11-year-olds. To work with younger children requires a new pedagogical approach which we call the discourse model. Generally speaking, its focus is the requirement that leaders must trust in advance in what the children or subordinates can do. In an educational context, trust in advance is a kind of presupposition that children are able to take responsibility, to care for others, to reason morally, to search for the best solution, to be altruistic and so forth (see Oser & Althof, 1993).

This fundamental presupposition can generate enormous power when it becomes basic to what happens in a Just Community meeting. If teachers believe that, eventually, they are in charge to decide, that such young children would not be able to jointly participate in decision-making responsibly, that the solution arrived at would not be the best one, then no Just Community can come about. The best analogy for what we mean when talking of a positive presupposition is that of a mother talking to her small child: she talks as if the child could fully understand, talk and react like a mature person, and this is the only power that really makes language development optimal.

3. A third element concerns the "discussion culture" of a school as a whole. The Just Community setting is not only a democratic decision-making institution, it is also a round table situation ("round table" in a symbolic sense) in which all participants must practise positive interactive communication. Rules are, for example, not to repeat what others already have said, not to hurt others, to listen to the opponents, to be open to arguments that contradict one's own position, to learn how to convince teachers and peers and how to resist pressure to subscribe to a certain opinion. This language culture is part of the interactional climate and part of the shared norms which are developed in the Just Community meetings and secured by a couple of second-order rules, such as the veto right, which states that specific subjects can interrupt the procedures at any moment if, for example, they sense that the dignity of participants or of other people is jeopardised.
4. The fourth element which has a specific character for the German model concerns the change of the school structure with respect to the possibility of realising a Just Community programme. Two elements are important: that all changes occur first bottom up and then top down, and that the limits of such a programme are well known.

Let us return to the main question: what are the black holes? What are the unfulfilled expectations which Kohlberg knew about but, nevertheless, was not able to realise? Where were the dormant ghosts *here?*

There are a couple of issues that I would like to discuss by referring to the European group of teachers and researchers mentioned above who changed the concept of Just Community in the way described. My statements of a more theoretical nature are based on (a) long-term experiences of creating and guiding these Just Community Schools and (b) on long-term datasets related to such schools which tell at least partly what we stand for.

Minimal morality

First there is a lack in the Just Community approach in its Kohlbergian form with respect to what we call "minimal moral goals". These minimal,

overlapping, moral goals are (a) to stimulate the next higher stage of moral development; (b) to sensitise children and adolescents to moral issues such as care, justice, truthfulness and tolerance, and to give occasions for the expression of moral feelings; (c) to support and enable moral acts and prosocial behaviour; (d) to establish a positive moral atmosphere in which "negative moral behaviour and risk taking"—necessary for learning, although not a goal in itself—can be experienced and an internal value system, combined with knowledge of the bad, can be developed; and (e) to give opportunities to build up ethical knowledge which acts as a point of reference for reflecting on one's own judgement and the judgements of others. Let us suppose that we would accept these goals altogether as a minimal agenda and let us then ask: what are the dangers and pitfalls of putting them into practice? How can we systematically deal with such complex and highly responsible and demanding issues within the idea of the Just Community?

In order to understand the term "minimal morality" (the expression stems originally from Adorno, 1976 and refers to a negative experience in a life-related, concrete context), I would like to present a story in which all these dimensions are represented. A foreman who had to supervise 12 craftsmen filled his pockets at work with different cables and screws to repair something at home. When he was about to leave the factory he saw the owner coming in. Startled, he turned back, hung up his jacket and pretended to do some more work. When he came back to take his jacket from the open locker the cables and screws had disappeared. He was scared. He feared the owner had emptied his pockets, which would have been a catastrophe. He regretted what he had done and he talked about it to his wife. She advised him to ignore the incident but to work hard in the future, which he did. Later he found out who had emptied his pockets (it was not the owner), and when he told us the story, he said that he had been angry and ashamed of himself.

We collected such stories in a research project in which we tried to relate real negative moral acts of adults to their judgements. What does the story tell us?

This prototypical situation addresses all five goals of minimal morality: namely (a) a cognitive disequilibrium, through the act of returning to work after the theft, having met the owner of the factory, together with a cognitive insecurity through the discussion of the act with his wife; (b) a feeling of shame and indignation about his own action; (c) an act of retribution: he wants to work better and harder in order to show that his act must be a reparation: (d) a feeling of responsibility for his co-workers: he should be a moral model, he cannot steal things from his own work place and be a thief in the eyes of the others; and finally (e) a brutal knowledge of the fact that if he had got caught he could never again be a foreman and that he must do penance in the eyes of the others.

Kohlberg never systematised such goals; he did not care much about the specific character of social–moral issues that were discussed in Just Communities. Often there was only a discussion in terms of (a), but (b) to (e) were

forgotten. Besides the work of the Child Development Project in California and Althof's 1995 Just Community, no such systematic effort has been made. Nevertheless Kohlberg knew about it, and I think his efforts to create a curriculum for New York High Schools with Cesar Previdi speaks for this. He collected several films, stories, dilemmas and picture materials (e.g. *The Lord of the Flies*). Another sign is that he tried to systematise the goals for the Just Community approach. A good example is provided by Kohlberg *et al*. (1975, p. 2). They mention establishing shared morals, democracy and a co-operative based on equality, the equal responsibility of the students, stabilisation of collective responsibility, building up a climate of trust, the development of a social contract and a constitution, a higher moral level of the group (as group), establishing teachers' authority based on the ability to mediate fairly in conflicts between students, and students and teachers, as well as stimulations and help concerning individual moral decisions and actions. However, these goals are not properly related to the criteria of minimal morality that I tried to spell out and they are not systematised. It was only a first will to begin with, the feeling of a necessity. Minimal goals have to be systematised and clarified—as my hypothesis—when it comes to a creation of a similar formulation of goals nowadays. This holds true especially for the content level; there has not yet been enough will to create a curriculum orientated to development. *Facing History and Ourselves* (Stern-Storm, 1980) has been an impressive attempt but only a small part of it, and Selman's film scripts were, similarly, partial contributions (Kohlberg *et al*., 1972; Selman & Kohlberg, 1976). Nor does it seem enough to collect a number of individual dilemmas as we did in North Rhine Westfalia (Reinhardt, 1990). The content, such as human prejudice, aggression, cultural integration (e.g. the *Voices of Care and Freedom* literacy project in the Boston area) and risk has to be systematically put together.

Global school reform and Just Community

Another missing piece in Kohlberg's Just Community is the integration of this approach into a general concept of school reform. I remember May 1984 when Ted Sizer and his staff came to Harvard, and Larry and Ted were thinking deeply and arguing about joining the Coalition for Essential Schools approach with the Just Community approach (as is now practised in the case of the Scarsdale Alternative School). This would have meant connecting the goals of moral and democratic education with at least the following five imperatives for better schools:

- give room to teachers and students to work and learn in their own, appropriate ways;
- insist that students clearly exhibit mastery of their school work;
- get the incentives right, for students and for teachers;

- focus the students' work on the use of their minds; and
- keep the structure simple and flexible (Prospectus of the Coalition of Essential Schools, December 1985, p. 1; cf. Sizer, 1984).

Thus, the Just Community must become embedded in a wider range of issues of "effectiveness and responsibility". Not only must the moral claim be addressed but also the social, intellectual, academic, aesthetic, self-control and school culture-related aspects. The Just Community—in the classical sense of participatory democracy with its focus on individual and collective moral learning—is too small a field of action, and must somehow become a broader basis for enhancing self-development in all these topics. Let me give five examples of issues discussed in the Just Community meetings of (a) a secondary school in North Rhine Westfalia, and (b) the Just Community in a primary school in Switzerland with which Wolfgang Althof now works, both as a researcher and as a practitioner.

Theme 1. Aggression and violence. Starting from an example students elect comrades who are responsible for intervening whenever violence is an issue (peer mediation).

Theme 2. Shaping and remodelling the school's playground with new and adequate possibilities for the children to play, with colourful paintings and so on.

Theme 3. One student refused to help a weaker learner in mathematics. This event is a reason for the Just Community to organise a whole system of academic support possibilities (other students help younger ones, there is an office for organising parents who help, every student in a classroom has a "godfather" who helps, etc.).

Theme 4. Organising international help—in collaboration with members of the community administration; for instance, organising a bus full of clothes or non-perishable food for war zones (which, currently, could be Bosnia).

Theme 5. The whole school decided how to start school on Monday mornings and how to finish it Friday afternoon. Each class has the task of doing something with the rest of the school, taking turns. For example, one class starts Monday with a theatre sketch, finishing Friday with an aperitif; others start Monday with a song and end Friday with a meditation.

Whereas the students in traditional Just Community schools mostly discuss justice matters, the issues handled in these more comprehensive school projects also concern the school culture in a broader sense: the ritualisation of school life, issues of the aesthetic arrangement of the school environment, issues of prosocial and helping behaviour, issues of academic pride and

work, and so on. This means that the rule setting and voting process of the Just Community is superseded and encompassed by life issues in general. Thus the Just Community approach is used for deciding more than justice claims.

The missing awareness of emotional conviction

Another important question, or a third "ghost", may concern the relationship between emotional commitment with respect to "moral community" and the aspect of explicit arguments on moral norms and rules for the school community. One has to ask seriously whether the theoretical background of the Just Community approach is orientated too much towards cognitive development, and whether the ethos of a community does not, rather, depend on what could be called emotional conviction. That does not mean constructing a dichotomy between the emotional and cognitive domain but, rather, to give special focus to the "moral heart" of a community. As soon as topics are discussed that affect real interests—conflicting interests of students and teachers and not only important, but very general moral issues—the power of the argument and the cognitive capacities may be largely overestimated. The ethos of a community does not mean that there must always be agreed upon by all people concerned—the "heart" and the "head" do not necessarily always go together. What is called the ethos of a community—in the original sense of the term—depends largely on rituals or implicit traditions, and—if necessary—situational subordination. A community cannot focus on "justice" and "reasoning" and good argument only—it also has to take seriously the ambivalent and often contradictory situations among the emotional states of mind of its members.

In *the Moral Life of Schools* (1993) Jackson *et al.* speak about "cultivating expressive awareness". The embeddedness of those expressive qualities is mediated not so much through arguments as through symbols. "The expressive meaning of all that we see and hear in classrooms expands and deepens as we go along, and it combines to do so for as long as we are willing to invest time and energy in probing its depths and stretching its boundaries" (p. 257).

However, we have to admit that there is not yet a clear rationale for understanding this expressive meaning in many facets of schooling. An example: after a Just Community meeting students were collecting signatures in order to cancel a decision that had just taken place. Many teachers and students who had worked for this decision were disappointed. They felt unfairly treated and they felt that in spite of a difficult exchange of opinions the dissent now seemed to be stronger than consent. How do we perceive the meaning of such an event? In most evaluations of Just Community meetings this dissent would be a negative sign in terms of the "moral climate". But it is not. Damasio mentioned in his book *Descartes' Error* (1994) that there

are only cognitive emotions in which the cognitive part is ordering the emotional and the emotional is giving life to the cognitive. And since humans can rarely think about the present but only about the future, this dissent and the collecting of students' signatures is a sign of emotional conviction that this school must be changed again and again.

In addition to this, we also do not yet have a real body of "emotional conviction" measures. However, some go in the right direction, and I would like to mention three of them. The first consists of Clark Power's Self-Esteem concept (1994) as a moral base. The dimensions of self-regard and self-respect on one hand, and intrinsic self-worth and self-acceptance on the other—as understood from a cognitive developmental perspective—lead to a better empirical estimation of the relationship between self-esteem and moral behaviour. Moral self-respect has something like an emotional conviction part. Nevertheless, two objections must be raised. The first is that this measure does not consider any communitarian aspect. The second criticism is that the stages do not include a systematic differentiation between moral feelings such as regret and shame and moral decision-making steps such as proceeding from good or bad acts to a change of mind, followed by retribution or other revised courses of action. There should be a relationship between minimal goals and moral self-esteem.

The second measure which approaches the community as a whole stems from Ann Higgins (1995). She developed a scale for measuring school culture. The dimensions of this scale are "normative expectations", "student-teacher/school relationships", "student relationships" and "educational opportunities". Many items of this instrument are related to emotional conviction, like "learning how to speak up and express options" or "students and teachers trust each other". Here, the criticism has to take a different direction. The scale has a lack of items that could express that a student cognitively does not agree to parts of, for example, a school discourse, is in dissent with certain rules and regulations but, still, emotionally, is positively related to the Just Community. That is why we argue that it is not always the power of the better argument that is the sustaining force but rituals, implicit traditions and the learning of an ethos of dissent, or an ethos of the difference (see Lyotard, 1986).

Battistich *et al.* (1995) successfully relate the sense of community (caring in classrooms, caring throughout the school, students' autonomy) to a wide range of attitudinal, motivational and behavioural variables and to social stratification. Their studies (that refer to interventional school community programmes) lack perhaps only a control of treatment validity: we do not know exactly how the emotional features that they measure really emerge. Nevertheless, many of their measures do regard the emotional part of community building as important.

We, the Fribourg group, are starting to develop a specific measure of moral self-efficacy belief. Bandura (1995) states that a "strong sense of

efficacy in socially valued pursuits is conducive to human attainment and well-being, it is not an unmixed blessing. The impact of personal efficacy on the nature and quality of life depends, of course, on the purposes to which it is put" (p. 1). Bandura himself did not develop a *moral* self-efficacy scale. This scale would measure how much, in critical situations, a subject believes that he has behaved successfully according to his/her own socio-moral standards. This scale would be an emotional conviction scale, and it could also be developed to measure collective school efficacy and even moral collective school efficacy (Oser, 1995). Bandura, in using this term (p. 20), is very vague about such a perspective. He speaks only about the prediction of the schools' level of academic achievement. There is no notion about shared norms, common values, lived dissent, school culture and so on.

All in all, I would like to show that cognitive emotions in a Just Community setting do not belong either only to the individual or only to the community. The co-construction of norms in the process of a Just Community debate contains an emotional or expressive awareness. There are presuppositions, trust in advance, positively lived dissents, hope for a better community, risk taking and so on. All cognitive emotions and norms are not generated as such, but only if they "destroy" other norms. Lothar Krappmann and I call this the "Orestes Effect". Orestes killed his mother because she killed her husband, his father, and he is only justified by the people, not accepting this death as such, but accepting it as a means for doing justice, if he himself does penitence. Co-construction does not mean "socially shared cognitions" or "shared norms" or "socially common knowledge"; it means developing an emotional conviction towards what is given up and what is new. Any regulation process in a Just Community meeting contains a destruction of a given state, a balancing of subjective and collective emotional commitment and an experience of moral self-efficacy. What we do not yet know is how the act is performed and how emotional conviction leads to execution of moral plans. Kohlberg's ghost here is his responsibility notion, his notion of shared values and his belief that there is an intrinsic benefit in what I do for the other.

Developmental emotional concern

With respect to the issue of cognitive emotions there is little fundamental research. There is one strand that I would like to discuss.

Asendorpf & Nunner-Winkler (1992, pp. 193–196) start to criticise the notion of a cognitive–affective parallelism that Kohlberg inherited from Piaget. They show that moral judgement and moral motivation are acquired in distinct learning processes. Nunner-Winkler and Sodian (1988) asked 200 4–8-year-old children how a protagonist who has been stealing feels after this action. Most of the younger children, even if they were clearly able to distinguish between moral and social realms (according to Turiel (1983))

attributed to the wrongdoers only positive feelings whereas older children saw that the wrongdoers will feel bad if they have had bad intentions. Younger children even think that a protagonist who resists the temptation to steal feels bad because of his "omission"; this attribution also occurs when the children know that to harm someone without intention leads to shame and compassion. Another interesting finding is that older children whose judgements are pre-conventional think that the protagonist feels bad not because of punishment but because of his unintended wrongdoing. If children in an experimental situation resisted cheating egoistically looking after their own needs in distributional situations, this was either due to shyness or to moral feelings. Moral behaviour where children act against their own rule understanding is due to their emotional attribution. It takes years until children learn to behave according to their own knowledge. Thus, Nunner-Winkler speaks about a two-phase model of moral learning; first there is the acquisition of knowledge regarding moral norms and only much later is the establishment of moral motivation (also see Asendorpf & Nunner-Winkler, 1992; Murgatroyd & Robinson, 1993).

Valid though this research may be, we do not yet know enough about the relationship between moral emotion (will or motivation) and respective stages. In one of his last papers, Kohlberg (1987) talks about conscience and morality, and he quotes his subject Joan: "What does 'conscience' mean to you? Well, I guess 'conscience' means the same thing as what I call 'moral responsibility"' (p. 13). And he adds: "I'd just like to conclude then in terms of our approach to the motivational basis for our approach to moral education. On the one hand, we have to develop the competence of moral judgment which we interpret primarily as a stage increase in judgments of justice. And second, to develop the capacity and motivation for moral action, which we view primarily as acting responsibly."

> There are two basic motivational principles here. The first is that, again, the student is capable of self-criticism. When there is a contradiction between his judgment what he's agreed upon and his action. There is a long process of democratic discussion about norms of rightness for each of the particular schools. If the student deviates from this norm which he has agreed to then he is subject not only to group criticism of course as having violated the agreements he has made with the community but he is subject also to self-criticism as having been irresponsible in his conduct. The second motivational principle (. . .) is really the notion that moral motivation involves membership in a moral community (pp. 14–15).

One can see that the relationship between responsibility and moral feelings is not yet elaborated. We do not yet know if the moral judgement development develops in advance if higher stages are concerned. Could we,

Table I Heuristic of the relationship between elements of cognitive and emotional development

	Emotional dimensions		
Stages according to Kohlberg	*Responsibility toward*	*Shame as*	*Indignation as*
0. Knowledge of rules	Own needs only	Not reacting to one's own needs	Hesitating to fulfill own need
1. Judgement according to punishment and reward Difference between physical and psychological	What belongs to me only	If one takes from others	Negative feeling because the wrongdoer does not reach his goal
2. Sense of justice based on rather simple notion of equality and concrete reciprocity	What is rewarded and not punished	If one is punished because one got caught	If someone else is punished because they got caught
3. Third party perspective, group feeling such as "care, love, trust which is mutually shared and mutually valuable" (Kohlberg, 1987, p. 8)	A reversible balance of giving and taking of equal share	If one does not care about the balance of giving and taking, of equal share	If someone else is treated unjustly because he did not receive what he believes he merits
4. The generalised other perspective. What if everyone would act the same way? Law, duties, and rights	A group feeling responsible for the welfare of others as a member of this group	If one does not behave according to shared norms	If other members of the group do behave against common goals and shared norms
5. Universal rights, the same rights without differences to everyone. Point of view beyond society. Principled thinking	Forward society as a whole with its structures, rights, and duties	If one damages good values and societal rights	If others do not care about political and societal testimonies

for example, hypothesise the following possibilities as emotional frames (Table I)?

Perhaps this heuristic is completely wrong, and perhaps we must, eventually, accept that the higher the stage the more parallelism between judgement

and emotion must be hypothesised. Thus, this is a heuristic that should be empirically tested. Nunner-Winkler proposed an important idea with her theory of postponed emotional reactions—this idea is part of this concept. It seems to make sense to argue that our moral judgement always precedes moral motivation, and that it defines the specific nature of feelings of shame and indignation. There is much work to do in this direction.

Implicit morality

In a Just Community meeting many implicit moral judgements can be heard. We often think that they are stronger than any other form of rational argument. What is an implicit moral judgement? Moral justifications are mostly treated as being under conscious control. However, there is some evidence that in situations where time is short, or in specific contexts, implicit or unconscious fashions of moral reactions occur. According to Greenwald & Banaji (1995), the identifying feature of such implicit cognitions is previous experience. However, the individual does not seem to know about this relationship between past experiences and the present situation. In other words, there is some kind of powerful belief that seems to tell us: "My moral intuition is right, and even though I am not able to justify my view and moral reaction at the moment or in this concrete situation, I am absolutely sure that I am right". Kohlberg mentions this phenomenon in his dissertation. Here, he states: "The mixture of a feeling of legitimacy to our own group-organized and systematized preferences together with some tolerance of other groups seems usually to be connoted by the concept of 'values"' (Kohlberg, 1958, p. 112). Later, he talks about the "value set" and the "intellectual conformity set" which indicate the way in which people are convinced that their judgement is right and that the others should have the same judgement. Kohlberg writes: "In its full sense, a moral set implies an orientation toward a right decision, a decision which an ideally good and wise person could make (. . .). The solution must do justice both to what the self believes and feels, and yet meet the situation" (Kohlberg, 1958, p. 128). Without further problematisation, Kohlberg then goes an and relates such a form of legitimation to Stage Three.

Since then, intuitive moral convictions seem to have been a forgotten dimension of moral psychology. However, it is a construct similar to what we previously called self-efficacy belief; namely, the conviction of how much I am sure that my decision is right and good and will "pay off" in the long run.

In our research on the responsibility of teachers, we found what we named "single handed" (unilateral) decision-making strategy—a form of decision-making that occurs in antagonistic situations where people might be hurt, or treated in an unjust or careless way. This form of decision-making is interesting because many board members of schools or firms believe that this strategy for solving moral conflicts is the best or most adequate because

it refers to implicit moral feelings which are supposed to guide the protagonist in an unambiguous manner. However, single-handed decision-making is not immune to stereotypes (concerning gender or race, for instance), and there is the possibility that such judgements inhibit reflective ways of moral performance. It seems that implicit moral judgements are related to rather unconscious value systems as products of encapsulated experiences and feelings of the past and that they are implicitly transmitted from one generation to the next. In our first study of the moral judgement of apprentices (of ages 16–19 years) we tried to assess the stability and rigidity of such implicit beliefs and related them to the moral stage. We found that individuals who could be clearly located on either Stage Two or Three were much more stable (or rigid) with regard to their implicit beliefs than individuals on transitional Stage $2^1/_2$. Therefore we concluded that implicit moral patterns are tightly connected to a clear localisation on the moral stages. (In this respect, I agree with Greenwald & Banaji (1995, p. 5) who write: "To the extent that implicit cognition differs from self-reportable [conscious or explicit] cognition, direct measures—that is, measures that presume accurate introspection—are necessarily inadequate for its study. Rather, investigations of implicit cognitions require indirect measures, which neither inform the subject of what is being assessed nor request self report concerning it.")

In general, Kohlberg's theory once again shows a black hole here. This theory is perceived as a rather philosophical system with some empirical validation. Kohlberg himself was not too much interested in how actual people really judge, feel and act. There are a couple of signs that seem to support this rather surprising hypothesis. One of them is the difference between competence and performance. Kohlberg was mainly interested in people's competences. He made a strong point in pushing people towards a better or optimal reversibility in moral judgement. But how do people judge under time pressure, under social and political pressure, or within an adolescent group that has great influence on every peer? This kind of question was not really attractive to Kohlberg. It contains implicit morality.

Process morality

Let us look now at a further problem that Kohlberg recognised but no one seriously tried to solve. This pertinent problem (another Kohlbergian ghost) concerns the question of teachers' expertise (a) with respect to the professional ethos and (b) with respect to the ability to set conditions for an optimal moral education. Both aspects are very problematic in the sense that there are no real training possibilities in most systems of teacher education. However, the Fribourg group on moral education has developed some perspectives on both aspects.

We define ethos as a form of procedural morality in dealing with interpersonal conflicts of daily life in schools or families. Sometimes the conflicts

Table II Cross-tabulation of moral stages and types of decision-making (second interview situation; pretest; percentages) (Althof, 1995, p. 290)

Stage	*AV*	*SE*	*SH*	*D1*	*D2*	*A1**	*N*	*Stage totals percentage*
3	—	2	3	3	—	2	10	23.8
3(4)	—	—	5	8	2	—	15	35.7
4(3)	—	1	3	1	—	2	7	16.7
4	—	—	4	3	1	—	8	19.0
4(5)	—	—	—	1	—	1	2	4.8
Type totals	—	3	15	16	3	5	42	
Percentage	—	7.5	35.7	38.1	7.1	11.9		100.0

* AV = avoiding; SE = security seeking; SH = single-handed decision-making;
D1 = discourse 1; D2 = discourse 2; A1 = Ambiguous or unscorable ethos interviews.

are of a more organisational nature, but sometimes they affect the dignity of persons including, for example, the possibility that somebody gets hurt in one way or other. In these situations, teachers have to be familiar with what we call the "round-table model" of conflict prevention or solution. The basic idea is that when several people are involved in a conflictual situation it is necessary to get them together and make a common attempt to solve the conflict. Often, this requires that the teacher first interrupts the flow of events and creates a situation in which a dialogue is possible (partner discussion, classroom meeting, Just Community meeting, etc.). Secondly, the teacher must guide these discussions in a way that all claims and needs and vulnerabilities can really be "put on the table"—that is, talked about in an open and trustful atmosphere. Finally, the teacher has to support all attempts to find the morally best solution. This includes discussing and co-ordinating the moral values at stake, and considering options under the perspective of moral criteria such as justice, care and truthfulness.

There are different ways in which people co-ordinate these claims of justice, care and truthfulness: avoidance, delegation, single-handed decision-making and discourse, in which all participants are responsible for the best solution (Oser & Althof, 1993). The point that I would like to address here is the relationship between this process of morality and the stages of moral development. Can it be that the optimal forms, the discourse forms, are possible at any stages? Can it be that the understanding of discourse can be better or worse even at the higher stages, such as Stage Five?

However, Althof (1995) showed in one of our studies that there is no empirical relationship between the level of moral reasoning and discourse type (see Table II). Althof concludes:

> A look at Table II shows that the tendency of higher-stage moral reasoning to be connected with more discursive types of decision-making is anything but impressive. Again, there is reason for a cautious standpoint. If we had hypothesized that a higher moral judgement should globally be related to a higher type of discursivity, this would be disappointing. But this has not been our hypothesis. Interestingly, a further calculation of the consistency of argumentation in terms of comparable attitudes with regard to both ethos dilemmas (discursivity or non-discursivity in both cases) shows clear differences between consistent (363 WAS points) and inconsistent subjects (337 WAS points), not between those who argue consistently discursive or consistently non-discursive, respectively. Evidently, subjects who are not (yet) in a process of leaving stage 3 tend more than the higher stage subjects to let their courses of action be dictated by situational conditions. All in all, our contention has at least not been disconfirmed that a high professional ethos needs at least a stabilized stage 3 moral judgment competence and that thereafter many other factors, aside from justice reasoning, enter the arena and decide whether more or less mature forms of professional morality are in use (p. 290).

Nevertheless, since teachers on any adult stage of moral development can learn to use complete discourse procedures there is no need to expect the discourse orientation on a higher stage only. In our interventional studies, we were able to show that teachers are able to acquire discursive orientations within three months (achievements that should not be due to minor increases in the moral reasoning scores within the same period of time).

To conclude, I would like to present the necessary steps that stand for a high discourse orientation, independent of the level of moral development:

1. Interruption of the situation and creation of a round table discussion.
2. Controversial exchange of information and perceptions about the given violation of rules and needs; the teacher's task is both participation and organization.
3. Presupposition of:
   (a) the ability to balance claims on an individual level;
   (b) the ability to coordinate the balance between the people or parties affected.
4. Practising the—sometimes—contrafactual belief that the solutions to be found are the best ones ("trust in advance").

These steps are of great importance in order to realise a complete discourse. They can be viewed as the horizontal complement of the task of climbing vertically, the stages of moral development. Kohlberg was confronted with

this whole stack when he met Jürgen Habermas in the 1970s; but he never developed the discourse idea further, as Habermas never cared about stages lower than Stage Five.

Teachers' behaviour in Just Community meetings

Many theorists in the field of moral education argue that there is much inadequate theoretical thinking among teachers. The most important mistake is, perhaps, that teachers do not believe in developmental measures. Every scientist knows that stage change measurement is a heavy and critical topic. It is easier to measure knowledge than to measure structural transformation. If they do not believe in the efficacy of higher stage thinking (autonomy, reversibility, etc.) they also do not believe in the necessity of interventions.

After Kohlberg's paper "Development as the aim of education", many teachers believed that the only way to instruct in a psychologically proper way is the dilemma discussion. It was a new approach at the time that provided (a) a clear psychological theory, (b) a transformation model (the subject's thought structure has to be disequilibrated, and new elements have to be build into a new structure), (c) an application technique for the practical teaching field and (d) a measurement possibility.

However, imagine a teacher who, after a well-constructed lesson in mathematics, experiences excellent results on the part of his students. Imagine how convinced he is that his teaching is important and how highly he would score in a professional self-efficacy measure. Now think of a teacher who participates in a Just Community meeting. There is no immediate result, there is no immediate higher stage or higher competence in conflict solving, no higher amount of capacity for argument and the like. One can only talk about processes, and processes lose their significance when one tries to control them continually. So it can happen that the meeting does not work well—meaning that no high professional self-efficacy is experienced and no such things as immediate success can be shown.

This situation has its aggravating special features if teachers are not well enough informed with respect to the theory of moral development, the concept of shared norms, the judgement-action problem, the validity of intervention, the stage transformation concept, the morality of care and so forth. In a research project on basic models of teaching, Oser *et al.* (1993, 1994) found that four-fifths of all teaching is knowledge or product orientated. Only a few intentional teaching interventions are focused on such issues as democracy, responsibility, value transformational development and developmental negotiation.

From this situation the question arises as to what teachers who participate in a Just Community programme do need—besides the discourse technique and the technique of developmental stimulation—in order to be able to experience success. Such needs become clearer when we consider practical

issues like those we discussed after the Just Community meetings in North Rhine Westfalia. For example:

- "Teachers didn't say anything; which makes the other participants somehow insecure (cognitive disequilibrium)."
- "Teachers taked too long."
- "Teachers did too much structuring of the meeting."
- "The Just Community was normative orientated: teachers preached."
- "Students didn't make any progress."
- "Students repeated the same standpoints too many times."
- "Students forgot the rules that were agreed upon earlier."
- "Students found creative solutions that the teachers did not think of before."
- "Students attacked the fairness committee."
- "All the teachers together didn't have a shared conception of values in this school."
- "Teachers inhibited a real controversy."
- "There is no progress in helping behavior."
- "We do not see progress in reasoning (stage transition)", etc.

Very often, after a Just Community meeting, teachers and students have the need to reconstruct and evaluate, discuss and analyse what happened in the meeting; but then the leaders must be prepared to take one issue out in order to treat it separately.

Hundreds of intervention studies give us information about (a) under which conditions stage transformation is possible; (b) how cognitive disequilibration is made possible; and (c) how the mechanisms of transformation can be understood. With regard to (a) we know, for instance, that the best influence on children is possible when they are confronted not only with artificial situations but with real-life dilemmas which are meaningful for the child and lead to a high involvement (Walker *et al.*, 1987). We know that stimulation is more effective in a respectful social context (see Lind, 1993); it is more effective for children in a transformational phase than for those who have just acquired a new stage of thinking. Intervention programmes work best when people are continuously confronted with dilemmas for at least three months (Schläfli *et al.*, 1985). With regard to (b) we know that adjacent stages with respect to two people discussing a moral dilemma are best for the stimulation of a disequilibrium (Rest *et al.*, 1969; Berkowitz *et al.*, 1980), that different forms of controversial discussions are helpful (e.g. all pros on one side, all cons on the other side of the classroom; group work; partner discussion; panel discussion, etc.). With regard to (c) we know that the main mechanism of stage transformation can be seen in the step-by-step erosion of the given stage structure (cognitive crises), the inclusion of new elements in the old structure and the new interpretation of the whole through these new elements (see Fig. 1).

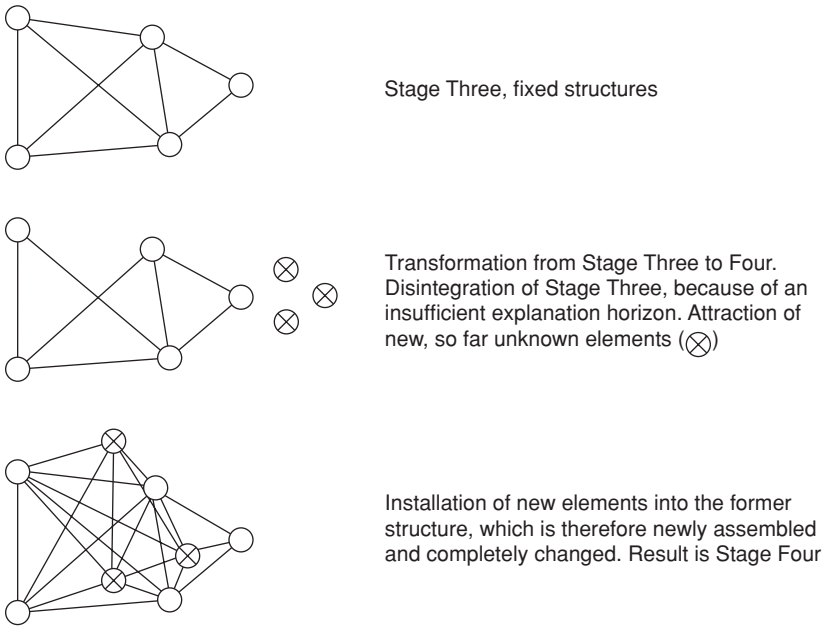


Figure 1 Transformation model as a basis for a development orientated discussion (example: Stage Three and Four)

But all this knowledge, fruitful as it may be, only helps us to foster a higher reversibility on the level of moral reasoning competence. Action priorities and capabilities are only slightly affected by discussion programmes. In saying this, I do not object to the cognitive–developmentalist position—certainly, moral judgement schemes are a deep structure of world interpretation. However, those structures play a role only on the level of interpretation and justification, and this is not enough. From history we know that the only thing in life that counts is the moral act. The process of arriving at a moral act includes a reflection about alternatives, a decision, a following-through of the act, a moral evaluation of the immediate result and of reflective consequences. From a psychological point of view, we know that it looks as though there is a so-called monotonic relationship between moral judgement and action. The notion of a monotonic relationship suggests that the higher the stage, the more solid and consistent the relationship between judgement and action. However, a closer analysis leads to the finding that this theory is insufficient or even false. Empirically, we are confronted with the fact that in most experiments concerning moral judgement and action, the higher stages are represented by a very small number of subjects. From a logical point of view, just because higher stages understand the complexity of problems, there is no reason why these subjects should have less difficulty of acting

according to their judgement than subjects on a low stage (cf. Oser, 1993). Higher stage citizens in Nazi Germany had the same difficulties in acting according to their judgement as lower stage people.

From an educational point of view, this means that no dilemma discussion and no stimulation toward a higher developmental stage should occur without relation to action and without contextualisation. Contextualisation means adequately considering aspects of moral performance. Contextualisation also means knowing how to create a Just Community and how to avoid all possible failures. The reference to action can take different forms. The more abstract form (a) consists in a reflection and discussion about the question "how would I really act in this situation?" I might say, in the case of the famous Heinz Dilemma (in which Heinz asks himself whether he should steal a life-saving drug for his wife) that, on one hand, Heinz should steal because life is a higher value than property. On the other hand I, myself, would possibly never dare to do so. Such inconsistencies tell us something about moral courage. The more concrete form (b) exists in real life dilemmas in which, for example, helping behaviour is demanded and enforced by the teacher. The most concrete form (c) is a decision of a whole community which is voted upon and enforced by this community. The Just Community approach to democratic schooling provides one form of a setting in which the act is constructed by all participants: students and teachers decide on the actual ways to solve conflicts, and they evaluate the decision-making outcomes commonly.

That is why a teacher must be familiar with more than a cognitive stimulation technique. He/she must be embedded in what I call participatory pedagogy.

Black holes are visions

In this paper, I have mentioned several educational issues that Kohlberg was aware of but never addressed systematically. I believe that the "black holes" referred to are ghosts that help to develop new research paradigms. Indeed, there is an implicit moral argument not yet investigated, there is an emotional argument, there is contextualised Just Community knowledge, there is procedural morality, and so forth. When Piaget worked with Binet in the early 1900s, he used all the mismatching cases of Binet's intelligence test to create a new theory. Perhaps we could also use Kohlberg's dormant ghosts in order to develop a contextualised theory of transformation which, in fact, would be a new educational theory. Education today is itself a dormant ghost. New visions are scarce. To embed the Just Community concept in a more comprehensive school reform with all the conflicts and tensions to solve, to use minimal morality as an interpretative frame might provide examples for a departure to new boundaries of meaning making. In this sense, education in a postmodern society can still create islands of hope; one of these islands could be the Just Community approach that introduces into school life a process towards a better founded moral literacy.

Acknowledgments

This is the text of the 8th Annual Kohlberg Memorial Lecture which was delivered at the 21st Annual Conference of the Association for Moral Education, New York, USA, 17 November 1995. The contributions of Wolfgang Althof to this revised version of the paper are gratefully acknowledged.

References

Adorno, T.W. (1976) *Minima Moralia. Reflexionen aus dem beschädigten Leben* (Frankfurt/M, Suhrkamp).

Althof, W. (1995) Teachers' moral judgment and interpersonal problem-solving in the classroom, in: R. Olechowski & G. Khan-Svik (Eds) *Experimental Research on Teaching and Learning*, pp. 271–294 (Frankfurt/M, Peter Lang).

Asendorpf, J.B. & Nunner-Winkler, G. (1992) Children's moral motive strength and temperamental inhibition reduce their immoral behavior in real moral conflicts, *Child Development*, 63, pp. 1223–1235.

Bandura, A. (Ed.) (1995) *Self-Efficacy in Changing Societies* (Cambridge, Cambridge University Press).

Battistich, V., Solomon, D., Dong-il., K., Watson, M. & Schaps. E. (1995) Schools as communities, poverty level of student populations, and students' attitudes, motives, and performance: a multilevel analysis, *American Educational Research Journal*, 32, p. 627–658.

Beck, K. (in preparation) Die Entwicklung von moralischer Kompetenz und beruflicher Leistungsfähigkeit als Problem der kaufmännischen Berufserziehung—Zur Analyse der Segmentiengshypothese. University of Mainz.

Berkowitz, M.W., Gibbs, J.C. & Broughton, J.M. (1980) The relation of moral stage disparity to developmental effects of peer dialogues, *Merrill-Palmer Quarterly*, 26, pp. 341–357.

Blasi, A. (1980) Bridging moral cognition and moral action: a critical review of the literature, *Psychological Bulletin*, 88, pp. 1–45.

Blasi, A. (1983) Moral cognition and moral action: a theoretical perspective, *Developmental Review*, 3, pp. 178–210.

Blatt, M. (1969) *The effects of classroom discussion programmes upon children's moral development*, unpublished doctoral dissertation. University of Chicago.

Blatt, M. & Kohlberg, L. (1975) The effect of classroom moral discussion upon children's level of moral judgement, *Journal of Moral Education*, 4, pp. 129–161.

Bruner, J.S. (1966) *Toward a Theory of Instruction* (Cambridge, MA, Harvard University Press).

Damasio, A.R. (1994) *Descartes' Error. Emotion, Reason and the Human Brain* (New York, Putnam).

Damon, W. (1977) *The Social World of the Child* (San Francisco, Jossey-Bass).

Dewey, J. (1944) *Democracy and Education* (New York, Macmillan).

Eckensberger, L.H. & Reinshagen, H. (1978) Eine alternative Interpretation von Kohlbergs Stufentheorie des moralischen Urteils, in: L.H. Eckensberger (Ed.) *Entwicklung des moralischen Urteilens. Theorie—Methoden—Praxis*, pp. 27–92 (Saarbrücken, Universität des Saarlandes).

Edelstein, W. & Keller, M. (1986) Beziehungsverständnis und moralische Reflexion. Eine entwicklungspsychologische Untersuchung, in: W. Edelstein & G. Nunner-Winkler (Eds) *Zur Bestimmung der Moral. Philosophische und sozialwissenschaftliche Beiträge zur Moralforschung*, pp. 321–346 (Frankfurt/M, Suhrkamp).

Fowler, J.W. (1981) *Stages of Faith. The Psychology of Human Development and the Quest for Meaning* (San Francisco, Harper & Row).

Greenwald, A.G. & Banaji, M.R. (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes, *Psychological Review*, 102, pp. 4–27.

Higgins, A. (1995) Dimensions of the school culture scale. Measuring attitudes, norms, and values in educational settings, unpublished manuscript (New York, Fordham University).

Jackson, P.W., Boostrom, R.E. & Hansen D.T. (1993) *The Moral Life of Schools* (San Francisco, Jossey-Bass).

Keller, M. (1984): Resolving conflicts in friendship: the development of moral understanding in everyday life, in: W. Kurtines & J. Gewirtz (Eds) *Morality, Moral Behavior, and Moral Development*, pp. 140–158 (New York, Wiley).

Keller, M. (1990) Zur Entwicklung moralischer Reflexion: Eine Kritik und Rekonzeptualisierung der Stufen des präkonventionellen moralischen Urteils in der Theorie von L. Kohlberg, in: M. Knopf & W. Schneider (Eds) *Entwicklung. Allgemeine Verläufe—Individuelle Unterschiede—Pädagogische Konsequenzen. Festschrift für Franz Emanuel Weinert*, pp. 19–44 (Göttigen, Hogrefe).

Kohlberg, L. (1958) The development of modes of moral thinking and choice in the years 10 to 16. University of Chicago, unpublished dissertation (Chicago, University of Chicago).

Kohlberg, L. (1987) Conscience as principled responsibility: on the philosophy of stage six, in: G. Zecha & P. Weingartner (Eds) *Conscience: an interdisciplinary view. Salzburg Colloquium on Ethics in the Sciences and Humanities*, pp. 3–15 (Dordrecht, Reidel).

Kohlberg, L. & Mayer, R. (1972) Development as the aim of education, *Harvard Educational Review*, 42, pp. 449–496.

Kohlberg, L., Selman, R.L. & Lickona, T. (1972) *First Things: values, filmstrips and discussion guides* (New York, Guidance Associates).

Kohlberg, L., Wasserman, E. & Richardson, N. (1975) The Just Community school: the theory and the Cambridge Cluster school experiment, in *Collected Papers on Moral Development and Moral Education*, Ch. 29 (Harvard University, Center for Moral Education).

Lempert, W. (1988) *Moralisches Denken. Seine Entwicklung jenseits der Kindheit und seine Beeinflußbarkeit in der Sekundarstufe II* (Essen, Neue Deutsche Schule).

Lind, G. (1984) Theorie und Validität des "Moralischen-Urteil-Tests". Zur Erfassung kognitiv-struktureller Effekte der Sozialisation, in: G. Framhein & J. Langer (Eds) *Student und Studium im internationalen Vergleich*, pp. 166–187 (Klagenturt: Kärtner Druck).

Lind, G. (1993) *Moral und Bildung. Zur Kritik von Kohlbergs Theorie der moralisch-kognitiven Entwicklung* (Heidelberg, Asanger).

Lind, G. & Link, L. (1986) Moralische Atmosphäre Fragebogen (MAF). Theoretische Grundlagen und Instrumente, unpublished manuscript (Universität Konstanz).

Lyotard, J.-F. (1986) *Das postmoderne Wissen* (Wien, Edition Passagen).

Murgatroyd, S.J. & Robinson, E.J. (1993) Children's judgements of emotion following moral transgressions. *International Journal of Behavioural Development*, 16, 93–111.

Nunner-Winkler, G. & Sodian, B. (1988) Children's understanding of moral emotions, *Child Development*, 59, pp. 1323–1338.

Nucci, L. (1981) Conceptions of personal issues: distinct from moral or societal concepts, *Child Development*, 52, pp. 114–121.

Oser, F. (1993) Die missachtete Freiheit moralischer Alternativen: Urteilen über Handeln, Handeln ohne Urteile, Berichte zur Erziehungswissenschaft, No. 95 (University of Fribourg, Dept. of Education).

Oser, F. (1995) Selbstwirksamkeit und Bildungsinstitution, in: W. Edelstein (Eds) *Entwicklungskrisen kompetent meistern*, pp. 63–73 (Heidelberg, Asanger).

Oser, F. & Althof, W. (1992) *Moralische Selbstbestimmung. Modelle der Entwicklung und Erziehung im Wertebereich. Ein Lehrbuch* (Stuttgart, Kletti-Cotta).

Oser, F. & Althof, W. (1993) Trust in advance: on the professional morality of teachers, *Journal of Moral Education*, 22, pp. 253–275.

Oser, F., Elsässer, T., Patry, J.-L. & Wagner, B. (1993) *Basic models of teaching-learning processes*, paper presented to the Annual Meeting of the American Educational Research Association, Atlanta, Georgia, 12–16 April 1993.

Oser, F. & Gmünder, P. (1991) Religious Judgement. A developmental approach (Birmingham, Alabama, Religious Education Press).

Oser, F. & Scarlett, W.G. (Eds) (1991) Religious development in childhood and adolescence, *New Directions for Child Development*, No. 52, San Francisco, Jossey-Bass.

Oser, F. & Patry, J.-L. (1994) Sichtstruktur und Basismodelle des Unterrichts: Über den Zusammenhang von Lehren und Lernen unter dem Gesichtspunk psychologischer Lernverläufe, in: R. Olechowski & B. Rollett (Eds) *Theorie und Praxis. Aspekte empirisch—pädagogischer Forschung—quantitative und qualitative Methoden*, pp. 138–168 (Frankfurt/M, Peter Lang).

Power, C., Higgins, A. & Kohlberg, L. (1989) *Lawrence Kohlberg's Approach to Moral Education* (New York, Columbia University Press).

Power, F.C., Khmelkov, V. & Self Esteem Project Team (1996) The Moral Basis of Self Esteem: a cognitive developmental approach (unpublished).

Reinhardt, S. (1990) Das Lehren muss konflikthaften Anforderungen gerecht werden, *Die Deutsche Schule*, 82, pp. 4–9.

Rest, J.R. & Narvaez, D. (Eds) (1994) *Moral Development in the Professions* (Hillsdale, NJ, Erlbaum).

Rest, J.R., Turiel, E. & Kohlberg, L. (1969) Level of moral development as a determinant of preference and comprehension of moral judgments made by others, *Journal of Personality*, 37, pp. 225–252.

Schläfli, A., Rest, J.R. & Thoma, S. (1985) Does moral education improve moral judgment? A meta-analysis of intervention studies, *Review of Educational Research*, 55, pp. 319–352.

Selman, R.L. (1980) *The Growth of Interpersonal Understanding. Developmental and Clinical Analyses* (New York, Academic Press).

Selman, R.L. & Kohlberg, L. (1976) *Relationships and Values. Filmstrips and Discussion Guides* (New York, Guidance Associates).

Sizer, T.R. (1984) *Horace's Compromise. The Dilemma of the American High School* (Boston, Houghton Mifflin).

Stern-Strom, M. (1980) Facing history and ourselves: integrating a holocaust unit into the curriculum, in: R.L. Mosher (Ed.) *Moral Education: a first generation of research and development*, pp. 213–233 (New York, Praeger).

Teo, T., Becker, G. & Edelstein, W. (1995) Variability in structured wholeness: context factors in L. Kohlberg's data on the development of moral judgment, *Merrill-Palmer Quarterly*, 41, pp. 381–393.

Turiel, E. (1983) *The Development of Social Knowledge, Morality and Convention* (Cambridge, Cambridge University Press).

Walker, L.J., DeVries, B. & Trevethan, S.D. (1987) Moral stages and moral orientations in real-life and hypothetical dilemmas, *Child Development*, 58, pp. 842–858.

Part XX

SPECIAL NEEDS

90

ARE WE TEACHING WHAT THEY NEED TO LEARN?

A critical analysis of the special school curriculum for students with mental retardation

J. W. L. Tse

Recently a group of Form 7 students wrote to a local newspaper complaining, perhaps misguidedly, about the lack of relevance in one of the subjects they took – music. Students with mental retardation would also have a lot to say about the issue of the relevance of learning materials, if they could freely and adequately express their thoughts in writing. While relevance in general education may not be a vital current concern, it is fundamental in the field of special education. Though students with mental retardation would be most unlikely to follow the footsteps of the Form 7 students, as teachers we must be held accountable and ask: Where are our students heading after leaving school? Are we helping them to develop their independence further, or are we just baby-sitting?

Most special education teachers would perhaps agree that the goal of education for students is to prepare them for a life as independent as possible in the least restrictive environment. A least restrictive environment is one that is as close as possible to that in which non-retarded people of the same chronological age usually perform. To teach skills which are of little significance is inconsistent with this goal. Though there are few curricular materials designed specifically for students with mental retardation, this does not justify providing them with curricula which are very often nothing more than a watered-down version of the regular school curriculum. Nonetheless, the skills that are most relevant are not always easily identified.

One way of studying curriculum content is to examine the aspect of time-on-task. The learning time which can be defined as "the amount of time students are judged to be engaged in learning content at a level of difficulty

Source: *A way of life: Living with mental retardation*, published by the Hong Kong Association for Scientific Study of Mental Handicap, 1989.

appropriate for them" is another additional guideline for teachers (Shavelson, Webb and Burstein, 1986). It seems appropriate to focus on the amount of time students devote to certain activities. This perspective concentrates on the student's level of functioning rather than in terms of the inherent characteristics of content. In special education, there is an apparent lack of research studies in the time-on-task perspective in general and curriculum evaluation in particular. When we examine the special school curriculum, questions inevitably follow: "Relevant to what?" "Functional to whom?" The purpose of this paper is to examine critically the curriculum for students with mental retardation in Hong Kong, and to suggest ways of building an ideal curriculum which hopefully can be implemented in the near future. Such an examination could have wider implications than simply for the territory of Hong Kong.

Severe mental retardation

Let us first examine some of the learning activities for students with severe mental retardation. The current curriculum (Special Education Division, Education Department, 1985) stresses four major areas or domains:

1. self-care
2. perceptual motor
3. social adaptation
4. communication skills

The following tasks, listed in the curriculum guidelines for students with severe mental retardation, are often referred to as cognitive, pre-academic or concerned with the concept of readiness.

1. Completing picture puzzles
2. Stringing beads
3. Stacking blocks
4. Looking at the person walking by
5. Concentrating on his own hand
6. Copying geometric forms
7. Using toy cups to drink
8. Discriminating colours and shapes
9. Arranging shapes of different sizes
10. Drawing lines

Mild mental retardation

Curriculum guidelines published by the Special Education Division, Education Department, in 1984 also focus on four main learning domains:

1. perceptual motor training
2. language
3. arithmetic
4. integrated studies

The following is a list of tasks in the curriculum for students with mild mental retardation which can be referred to as generic and nonfunctional:

1. Picking up things, digging, and rubbing
2. Stacking rings on a post
3. Walking on balance beam
4. Counting backwards
5. Multiplication and division
6. Fractions and ratio
7. Studying figure-ground relationship
8. Studying 200 basic words
9. Writing lines
10. Balancing something on the head and walking

The above-mentioned twenty tasks describe the pre-academic, readiness or cognitive skills listed in the curriculum guidelines. These should either be replaced by functional skills or they should be removed altogether.

Characteristics of the local curriculum

With the above examples in mind, it is not difficult to point out some of the major characteristics of the special school curriculum for students with mental retardation in Hong Kong which can be characterized as:

1. uniform
2. bottom-up (developmental-based)
3. norm-referenced
4. non-functional
5. generic skills
6. readiness-related
7. "the tail has wagged the dog" phenomenon

Uniform curriculum

Applying a uniform curriculum to the retarded population, which is so diverse in terms of intellectual and adaptive functioning levels, is bound to fail. An obvious reason for failure is that a uniform curriculum does not give adequate attention to individual needs and overlooks the specific ability and learning

style of each student. The condition is particularly serious when the IQ range for a level of mental retardation in Hong Kong may be as great as 25 points; the variability of skills in all areas is enormous. A common curriculum means that, although students with mental retardation are so diverse, they have to receive the same content materials. Nowadays, the selection of teaching areas and the training of these students can no longer be regarded as two separate processes but as essentially inter-related activities. However, the existing curriculum design inevitably leads teachers to focus more on what the students cannot do than on what they can do. The best way of assessing a student's ability in performing a task is by trying to teach him to do it. A more balanced use of testing is necessary, that is, less time spent testing and more time spent teaching. Assessment which stresses the nature of a deficit and seeks to remedy it should be replaced by the habilitative model which seeks to discover also the individual's assets in order to build on them.

Bottom-up (developmental-based)

There are few curricular materials designed specifically for students with mental retardation. Most curricula used with these students are referenced to models of normal human development which track the progressive refinement and elaboration of basic motor, social, and cognitive skills into complex and competent performances characteristic of normal adolescents. Such an approach directs parents and teachers to select goals with reference to the sequence of objectives in the developmental scale rather than in response to the practical performance needs of students or their preferences. These models of human development are typically stage theories that divide development into relatively discrete stages through which **all** children proceed en route to becoming independently functioning adults. The curriculum sequences derived from such stage theories can be described as "bottom-up" sequences. In applying this model, teachers try to identify those tasks or activities, normally performed by infants and young children, that the student has not yet mastered. Failed items are then treated as teaching objectives based upon the assumption that students with mental retardation should acquire the same sequence or progression of skills seen in children without such handicaps. They begin by teaching those skills which presumably occur first in a "normal" developmental sequence and then proceed to skills which occur at progressively older ages.

Since students with severe mental retardation manifest significant skill deficits accompanied by additional physical disabilities, they frequently receive instruction only on curriculum objectives characteristically offered to and mastered by infants or very young children. As they become adolescents or young adults, the outcome of such curricular strategies often results in the delivery of instructions which are not needed by the students or merely

serve to reinforce their dependency on others (Snell and Grigg, 1987). If the student has grown much beyond childhood, a developmentally-based curriculum is likely to have little impact upon the ultimate attainment of independence (Holvoet, Guess, Mulligan and Brown, 1980). Though they may in fact show evidence of progress through bottom-up curriculum sequences, the obvious performance discrepancies between them and their nonhandicapped chronological age-peers actually increase over time. A very real question confronting educators is related to how long to retain these students in bottom-up curricula. Given a limited number of years in school programmes, can the student possibly progress fast enough or far enough to acquire the skills needed for the most independent functioning possible in complex, heterogeneous post-school (hopefully, working) environments?

Norm-referenced

Norm-referenced assessments such as the AAMD Adaptive Behaviour Scale (Nihira, Foster, Shellhaas, and Leland, 1974), Vineland Adaptive Behaviour Scales (Sparrow, Balla, and Cicchetti, 1984) and a Hong Kong Based Adaptive Behaviour Scale (Kwok, Shek, Tse and Chan, 1989) are used in the identification process to verify the results of screening and to determine the extent of intellectual deviation from the norm. Such evaluation procedures are extremely useful in the classification of intelligence and identification of special abilities. They also provide a global picture of an unfamiliar student's abilities and provide common ground for reporting performance levels known to teachers and psychologists. However, these tests, at present, tend to remove testing from teaching. The focus is on the discrepancies between normal and observed student performance, with the consequent identification of "deficits" in student performance as priorities for intervention. While such discrepancies may certainly be viewed as deficits, it does not necessarily follow that the behaviours exhibited by normally developing children should be curricular priorities for every student with mental retardation.

It is often difficult to know what a score means or how it can inform actual teaching activities and it often does little to guide teaching and curriculum content. Ideally a test should tell us something about the strengths and weaknesses of a student's thinking and behaviour. Competent teachers are able to change their instructional strategies in order to help students surmount a difficulty that the assessment has highlighted. But are we able to do this for more general reasoning and thinking? We are able to discover that a particular student has an IQ of, say, 65. But can we say just what is making his/her score lower than the others, and can we use this information to try to teach him more efficient ways of reasoning and thinking? Typically, norm-referenced tests provide limited information about the ways students with mental retardation learn or about ways to ameliorate their handicaps.

Since norm-referenced assessment devices are not intended to detect small changes in behaviour and do not address all areas where instruction is likely, they cannot be used to evaluate student performance. Much has yet to be learned about tailoring remedial strategies to individual patterns of learning ability. Unfortunately in the process of finding out what to teach, teachers may rely too heavily on norm-referenced materials.

Nonfunctional

Nonfunctional skills are those that have an extremely low probability of being required in daily activities. Balance training such as walking with something balanced on the head and arithmetic skills like counting numbers backwards are **nonfunctional**, since the probability of using such skills is remote. Furthermore, such skills do not enhance the ability to live independently. Such skills are also rarely performed by nonhandicapped adolescents and adult members of society. Walking on a balance beam and counting numbers backwards are only functional if we are giving specific training to fashion models or professionals engaged in the aerospace industry. On the other hand, mere pencil-and-paper tasks can be easily arranged in class, but they are useless unless the students can apply these acquired skills in real-life situations. One of the class assignments, for example, in a school for students with moderate mental retardation, was addressing envelopes. When most instructional time is in trying to teach "academic" skills, little or no time is available for functional skills which students will require in the running of their everyday lives. It is quite legitimate to question the instructional validity of the skills we attempt to teach.

Teachers might teach students to zip a zipper on a zipper-board, but this is nonfunctional. Teaching students to zip his/her own pants is functional. Some well-intentioned teachers even go as far as to put a candy in the zipper-board as a reinforcer, but what behaviour is it reinforcing? In reality, most people zip or unzip for entirely different and quite automatic, functional reasons. Other traditional nonfunctional training activities may, for example, include the writing of certain Chinese characters, even though the students cannot read or identify real characters (reading and writing is no different in western countries; the "23 year old" Janet still throws the ball to the "27 year old" John!). Or the stacking of rings on a post is laboriously taught, even though the students are unable to stack dishes properly. Walking on balance beams, uttering pre-verbal sounds, and sorting colour chips are nonfunctional skills which have relatively little importance. The most significant needs of students with severe mental retardation is to learn to be more independent, for instance, by mastering **basic** self-care skills, such as bathing, self-feeding, toiletting, shopping, etc. Though bathing and taking a shower are included in the curriculum guidelines for students with severe mental retardation in Hong Kong and in other countries too, it is less common to

find special schools with normal bathroom facilities. Lunch hours are for teachers to take a break while some ancillaries take over and virtually, and in many respects actually, spoon-feed the students. The teaching of functional skills, especially in vocational areas like how to seek a job, is noticeably lacking.

If we are to train students to be more independent and to hold reasonably responsible jobs, then the components of vocational training are sadly lacking in the special school curriculum. Even if students with mental retardation can find jobs, research studies have consistently shown that they lack the social skills required in that type of setting. These studies have shown that employees could not socialize with others and more often than not have to quit their jobs, not because of lack of skill, but because of poor social relationships. Unfortunately the importance of developing appropriate social behaviours in the work setting has been grossly underestimated.

Generic skills

Examples of generic skills may include finger dexterity (putting pegs in a pegboard); colour discrimination and eye-hand coordination (stringing beads). For students with mild mental retardation the materials for instruction may consist of copying geometric figures, using toy telephones, or discriminating between shapes, e.g., triangles and squares. These activities are nothing more than **generic** skills. Improvement in the performance of such tasks rarely generalizes to improvements in the performance of other tasks requiring the same kind of skills (Hammill, 1982). In fact, failure to generalize learning from the classroom to the natural environment is perhaps the major problem facing students with mental retardation and their teachers.

Readiness-related

Readiness is generally referred to as a state of preparedness for more complex learning. Drawing lines, concentrating on one's own hand, using toy cups to drink from and other activities related to pre-verbal training are examples of the kind of teaching which primarily focus on the notion of "readiness". It is important to point out that learning "readiness" skills may not result in improved performance or learning of **functional** skills or independent activities. Some teachers claim that pre-verbal, academic skills are essential for progress in literacy and numeracy. Programmes directed at developing specific cognitive skills have been suggested as having significant transfer value to basic attainments. Others have explained that visual motor or visual perception programmes fulfill a valuable function in preparing students for progress in reading. In fact, such characteristics are often key features of reading readiness programmes. Research studies,

however, have largely rejected this view (Myers and Hammill, 1976). We are thus left with a question: how suitable are average programmes directed at developing pre-academic skills in the curriculum? Often students never learn to read, to write or to do arithmetic well enough to use these skills in any kind of functional manner. In the case of students with mild mental retardation, it could well be said that such programmes are totally inadequate. Emphasis on motivation and interest rather than on discrete cognitive abilities may be more preferable. In most cases, students being taught "readiness" or pre-academic skills are required to learn one set of skills that are **not** functional, then to learn a second set of skills that **are** functional. Naturally, valuable time is wasted. The end result is to catch the students in a trap where they are perpetually "getting ready", but are rarely doing anything of value in the community. In other words, we are constantly waiting about hopefully.

"The tail has wagged the dog" phenomenon

In the curriculum there is an implicit view that what is to be taught depends more on how students with mental retardation learn (i.e., process) and how teachers teach (i.e., methodology) than on what kind of learning task is important to the students. A concentration on students' learning processes has appeared to determine, at least in part, the content of the curriculum (Thomas, 1985). An interesting example is modelling others' behaviour. Is it a process, methodology or content? The significance attached to visual motor activities serves as another familiar example. It reflects the strong links which special education has had in the past with the discipline of psychology, especially with the psychology of learning: the focus is naturally placed on the **process** of learning. To pay close and frequent attention to how students learn invites teachers to consider questions of **how** to teach rather than **what** to teach. There seems to be a blurring of the edges between learning activities (content) on the one hand and learning characteristics (process) and teaching strategies (methodology) on the other. A focus on process soon becomes a question of determining what cognitive processes or learning patterns are important, redefining them as "abilities" or "skills" and drawing up a programme to develop them, regardless of whether these are appropriate or not. Hence, we have visual perception programmes; auditcry perception programmes; programmes for developing fine motor skills, gross motor skills, language skills, and so on.

In sum, parents and teachers tend to under-estimate the abilities of students with mental retardation. One thing we have to bear in mind is that the curriculum is designed for students with mild mental retardation from the age of 6 to 16. Sometimes one may wonder: Is mental retardation related to a failure of learning or a failure of teaching? Is there a chance that we are teaching them to be retarded? In the curriculum guidelines great

emphasis has been placed on safety: general knowledge about safety, road safety, work safety, safety in using tools and machinery, safety at home. Remember, as Perske (1972) wisely pointed out, there is also "dignity of risk". If we are over-protective, we are in fact limiting the learning opportunities of the students and degrading the integrity and dignity of these individuals. Calculated risk is a fact of life. The question is: Are we prepared to take it? We who are supposedly "normal" do this every time we get behind a steering wheel of a car.

Ideal curriculum

In simple terms, an ideal curriculum must be effective in bringing about significant progress in all meaningful areas for students with mental retardation. The norm-referenced type of curriculum which is offered to students with severe mental retardation is inappropriate. The curriculum should stress the importance of chronological-age-appropriate functional skills in natural environments. An ideal curriculum for students with mental retardation has to incorporate all of the following characteristics:

1. individualized
2. criterion-referenced, assessment-oriented
3. functional
4. chronological-age-appropriate
5. task-specific
6. natural
7. ecologically valid

Individualized

A universal phenomenon that characterizes all human beings is that human behaviour is unique. In fact, students with mental retardation are exceptional with highly individualistic needs and varied abilities. Without any form of tailor-made curriculum, individual characteristics have to be compromised, if not deliberately ignored. Mass education, on the other hand, requires that all students in a class learn the same things at the same speed in response to the same instruction. Individualized education means that the pace and method of instruction should be matched to each student's learning ability. Perhaps more importantly individualized education matches the content of instruction to each student's needs. Emphasis is placed on individualized and student controlled types of learning for this holds particular promise for students with mental retardation. This was recognised by Seguin as early as 1846. Since students with mental retardation could not be expected to learn **all** that other students were learning, and certainly not at a uniform speed and under uniform instruction and under similar circumstances,

individualizing was clearly more desirable. Ideally, everyone including parents, siblings and others concerned with the training and care of the student should be involved in planning an individualized education programme. Such an individualized curriculum which takes care of the individual characteristics of the student has to be designed especially for every student. In fact, in some countries such as the US each student with mental retardation is legally entitled to have his/her own individualized education plan. This is even more essential today than it was in the year 1846.

Criterion-referenced, curriculum-based

As stated earlier, norm-referenced measures are commonly used to certify the individual for programme eligibility: such tests tend to have limited instructional relevance. Standardized tests, such as Stanford-Binet Intelligence Scale (Thorndike, 1973) and Vineland Adaptive Behaviour Scale (Sparrow, Balla, & Cicchetti, 1984), need to be supplemented with new tests and techniques which are criterion-referenced, together with other informal assessments that are designed for specific assessments. Criterion-referenced measures are used primarily to determine the extent of deviation from mastery level – in other words, skill levels necessary for independent functioning: these measures form the core for assessment within the classroom. Examples of such measures are Balthazar Scales of Adaptive Behavior (Balthazar, 1976), Pennsylvania Training Model: Individual Assessment Guide (Somerton-Fair and Turner, 1979), Basic Life Skills Scale (Dalton, Cibiri, Baker, Malik and Wu, 1981) and Bereweeke Skill-teaching system (Dell, Felce, Flight, Jenkins & Mansell, 1986). Criterion-referenced measures have high instructional relevance since items are selected from a task analysis of skills taught in school. Criterion-referenced measures can help the teaching process directly. The concern is not with relative performance (how an individual compares with the rest of his age group) but rather with finding out whether or not the individual is able to do a defined task. Skills in a specific area may be arranged in sequential order, increasing in complexity. Each point can only be successfully accomplished if earlier and easier items have been mastered. Thus, in order to add up numbers, one has to be competent in 'easier' skills. Ideally, criterion-referenced tests provide information as to the level of skill competence acquired and should inform the teacher where to start teaching. Scores provide explicit information as to what an individual can/cannot do. Teachers and other direct care staff are encouraged to use criterion-referenced measures as assessment tools for collecting helpful information on the individual in various content areas, including gross motor, fine motor, cognitive skills, communication, social and self-care behaviours. If such measures are not available, teachers can use informal assessment, that is, constructing tasks in the above areas for the individual to perform, but ones appropriate to that individual's needs.

For any curriculum to be effective, it has primarily to be competency-based – that is, training would start only after a thorough and comprehensive assessment evaluation had been performed on a particular student. Assessment and training of students with mental retardation can no longer be regarded as two separate processes but as essentially inter-related activities. Assessment must be linked with instruction (Crawford and Tse, 1988). Where do we begin? or What do we begin with? Teachers can understand the student only when they know more about the strengths and weaknesses related to the assessment, and this should form the basis of the curriculum materials.

Functional skills

Functional skills refer to the variety of skills that are frequently demanded in natural domestic, vocational, and community environments (Brown, Hamre-Nietupski, Pumpian, Certo, and Gruenewald, 1979). Functional behaviours are skills of immediate usefulness, given a normal environment, and the training of which employs materials that are real for those environments, rather than simulated (Gannon, 1986). Functional skills are not limited to survival skills or the physical well-being of an individual; they may also include the variety of skills which influence the individual's ability to perform as independently and as productively as possible in home, school, and community. Nonfunctional skills are those that have an extremely low probability of being required in daily activities. Fitting pegs into a pegboard is relatively nonfunctional (unless you are training the individual to play games such as cribbage). Skills used in identifying exit signs, restroom signs, and vending machines are more functional than many other skills which occupy teachers' time.

An important aspect of functionality is where the individual will be in future (Brown, Nietupski, and Hamre-Nietupski, 1976), let us say in five or ten years time. The identification of future environments has direct implications for the prioritization of particular skills within curriculum domains. For example, if an adolescent's projected sheltered employment is connected by tram, bus or train, then school experience should develop the appropriate skills required to use the necessary transportation. A word of caution is called for here: because of differences between individuals, skills that are functional for one student cannot be assumed to be equally so for another. Instructional objectives must be determined individually: what appears to be useful may not be so for each person.

Chronological-age-appropriate skills

If one of the goals of education is to minimize the stigmatizing discrepancies between students with mental retardation and their nonhandicapped peers,

it is our obligation to teach the former the major functions characteristic of their chronological age using materials and tasks which do not highlight the deficiencies in their repertoires. For example, since young children typically play with simple wooden puzzles as a recreational activity, it is quite inappropriate to teach older children with severe mental retardation to play with such puzzles. Since nonhandicapped 20-year-olds are unlikely to spend recreational time assembling a four-piece puzzle of Donald Duck, it is stigmatizing to teach adolescents with mental retardation to do this. Either a different leisure activity (e.g., listening to records, engaging in a craft) or solving a puzzle with more age-appropriate content would be preferable. It appears that the selection of appropriate training activities can have a significant impact on the manner in which the individual are perceived by others in the community.

Since nonretarded adolescents and young adult students make independent purchases at market places, department stores, and drugstores, shopping skills should receive considerable attention in the curriculum for students with severe mental retardation. Even though teaching may be relatively prolonged, it is imperative that the students receive instruction in such major life skills.

Task-specific

Instead of generic or universal skills, task-specific activities should be encouraged. That is, any task taught should be direct. For example, if the instructional objective is to teach the student to identify bus number 46, writing 1 to 50 would not help much in enhancing effective learning. Regardless of teaching activities, the materials to be presented should be programme specific; they should be derived from the specific goals of the educational programming. Direct instruction of the specific task or clusters of related skills eliminates the problem of generalization. Very often teachers can make use of the school setting in order to teach a number of task-specific and generalization skills: for example a variety of bathrooms (self-care and janitorial skills); sidewalks/pavements (pedestrian skills and social behaviours); and hallways (social interaction and mobility). Skills should be taught directly in the form and settings where they must eventually be performed, thereby addressing directly the problems associated with the generalization of classroom instruction (Stokes and Baer, 1977).

Natural

Students with mental retardation do not learn skills easily or transfer them to new situations (Liberty, Haring, and Martin, 1981). If they are taught functional skills, using real materials (pay telephone, groceries in a store) rather than simulated materials (toy telephone, pictured groceries), they are more likely to be able to perform the skill outside the classroom. Natural

environments refer to the variety of least restrictive environments in which students with mental retardation function. These environments are important to curriculum development both as a location for training and as a source of curricular content. Artificial or simulated environments such as crossing the road in the hall and grocery shopping in the classroom store are organised out of wishful thinking with little chance of generalizing into real life settings. When the training setting is natural and the instructional focus is on task-specific skills, the problem of generalization can be easily tackled. When teaching a student to cross streets, an appropriate strategy is to start instruction on a side street and then gradually move to a busier street as the student masters the skill. Of course, simulated training in the classroom is perfectly safe: skill generalization is likely to be more difficult since the classroom differs drastically from a real busy street. In addition, it is just a bit difficult to simulate Nathan Road (Hong Kong) or Oxford Street (London) in the classroom, or even in the hallway. In reality, the actual use of pedestrian skills "learned" under simulated conditions may ultimately be more dangerous (Snell and Zirpoli, 1987). There are far more variables in reality than can ever be simulated in the classroom.

Ecological validity

The ecological approach asserts that all behaviours may be seen as occurring within some environmental context. The interaction between the behaviour of an individual and the various environmental contexts in which that behaviour occurs may be defined as the individual's behavioural ecology. It is important to stress not just the behaviour of a person, or the single environment in which it occurs, but rather the interaction of each with the other (Barker, 1968).

In discussing the ecological validity of experimental studies, Brooks and Baumeister (1977) suggested that the skills trained in the research studies should have some validity in the everyday experience of the subjects. They urged researchers to provide this rationale for their measures and show the real life basis on which their subjects differ. Why not constitute subjects on the basis of skills clearly important for adaptation – academic skills, language skills, motor skills, social skills – rather than etiology or IQ? Why not target for analysis the tasks that are relevant to the subjects' adaptation to their environment rather than the contrived arbitrary measures so popular in laboratories (e.g., discrimination of objects, and marble dropping)? Brooks and Baumeister urged researchers to carry out studies in the context of environments which are natural to the population and tasks under study.

The need for ecological validity cannot be overstressed as it is more liable to lead to skills to solve the problems of training and educating students

with mental retardation in community settings. Many teaching targets such as stacking blocks present a much easier object of training than a behaviour such as using a lift or escalator, which is essential for survival in everyday complex communities. Educational programmes for adolescents and young adults with severe mental retardation should be focused on preparation to function as independently and as productively as possible in non-school and post-school environments. Thus, it is suggested that the teaching of skills that are only appropriate in school environments should be minimized, and the teaching of skills that are appropriate in nonschool and post-school environments should be maximized. The selection of any teaching targets for students with mental retardation should emphasize ecological relevance. An ecologically valid task would:

1. target behaviours for change that are relevant for the adaptation by the students to their normal environment. That is, the teaching targets would be functional for the students rather than chosen on the basis of convenience to the teacher.
2. train in, or for, settings which are natural for people with mental retardation. This means under normalization settings (everyday complex communities) rather than restrictive settings (institutions).

A curriculum must be viewed in terms of its ecological validity: does it provide the principles and skills which students will incorporate into their behaviour repertoire and in turn use to solve their problems and meet their needs as adults? The ultimate worth of curricula for students with mental retardation is severely limited if it cannot provide workable procedures to assist them in achieving and maintaining adaptive skills in normal community environments. The fact is that even people with severe mental retardation are capable of living and are now seeking to live in these less restrictive environments. Teachers of students with mild mental retardation should be held accountable for failing to assist students to seek employment and to live in normal environments. If an educational community is going to have any real impact on the degree to which normalization will succeed, then a greater effort needs to be made in ensuring that the curriculum for students is ecologically valid. The lack of attention to social validity is scandalous, considering that society is ultimately the source of funding of the educational institutions.

Selecting relevant teaching areas

In the focus of assessment, there has been a shift from a developmental model to one which stresses practical life skill experiences in a person's present and future environment. The logic that traditional developmental sequences will necessarily lead to the acquisition of functional life skills

is questionable. A valid curriculum must emphasize age-appropriate functional skills demonstrated in natural community settings. Such a functional curriculum requires teachers to re-examine their methods of assessment and selection of training activities, and this is especially so for students with moderate and severe mental retardation. Apparently, the most appropriate action is to build an **ecological inventory**, which is to list skills necessary for functioning in five major domains: vocational, recreation/leisure, domestic, school and community as they occur in home, school, and local community activities (Mattison and Rosenberg, 1985).

An ecological inventory is a very useful decision making assessment tool for approaching the problem of "what to teach" students with moderate and severe mental retardation. The steps in building one are as follows (Brown et al. 1979):

1. Identify a target group: The target group is defined in terms of age and level of functioning. For example, separate ecological inventories would be required for:
   a) children aged one month to three years; four to six years, etc.;
   b) students with mild or severe mental retardation.
2. Identify curriculum areas: A survey should be made of the home, community, school and vocational environments in which each student lives.
3. Divide the environments into a variety of natural sub-settings in which students with mental retardation function or might be expected to function. A home, for example, can be divided into bedroom, bathroom, kitchen and living room: skills necessary for adaptive functioning in each of these sub-settings can then be identified.
4. List the activities that are mandatory for successful adaptation in the sub-settings for students. Important activities for a bathroom sub-setting include showering, tooth-brushing, toileting, cleaning the floor etc.
5. Assess student's performance of these sequences of behaviour and then assess progress made.
6. Design and implement teaching programmes for the identified skills.

Thus, by comparing the students' current skills with those skills identified as being necessary for success in future adult settings, an ecological curriculum could provide a truly individualized set of adaptive targets for instructional activity. This top-down approach will generate the identification of a large number of functional activities and skills. However, instruction could not possibly begin simultaneously on all skills and be equally effective. In selecting priority teaching tasks, teachers can rank the relative importance of different tasks. They can avoid imposing personal preferences in the selection of teaching areas by adopting the Task Importance Rating Scale (Baine, 1986):

Task Importance Rating Scale

The importance of each task may be judged by its contribution to:

	Low	*Medium*	*High*
1. understanding functional skills	0	1	2
2. increasing social competency	0	1	2
3. acquiring chronological age-appropriate skills	0	1	2
4. knowing survival skills	0	1	2
5. increasing opportunities to interact with non-retarded people	0	1	2
6. increasing ability to accomplish frequent task demands	0	1	2
7. improving health	0	1	2
8. broadening opportunities to understand/express ideas, and feelings	0	1	2
9. increasing opportunities to enjoy social recreational life	0	1	2

To use the Task Importance Rating Scale, parents and teachers should independently rate the relative importance of each task by giving an overall rating of low, medium or high importance in relation to other tasks the student may be required to learn. Tasks that are rated as relatively unimportant are eliminated while the task with the highest number may be the most important task for the student to learn.

Let the student with mental retardation choose

If at all possible, students should be encouraged to **choose what to learn**, what behaviour to change, or what to "earn" for performance of a skill. The individual's preferences should be considered when teachers are developing the teaching plan especially for activities (e.g., leisure) that typically provide choices. Assessment of the student's preferences for activities to learn or to earn could be signaled through his or her language system (e.g., pictures). Teachers can provide students with limited exposure to novel activities which can later be used as reinforcers. Sampling need not be limited to reinforcer selection but may also be used to assess the individual's preferences for new materials or activities. Providing choice can also be incorporated in lessons by providing alternative learning tasks or reinforcers. The ethical benefits of self-determined learning contents are evident as the student makes decisions regarding his or her education.

Limitations

Changing the curriculum is only one of the ways of improving the education and training of students with mental retardation. It is unfair to point out only the weaknesses in the curriculum without also mentioning some other areas where improvements could be made.

1. Improving student/teacher ratio

The student/teacher ratio is 20:1, 10:1, and 8:1 for mild, moderate and severe mental retardation respectively in Hong Kong, and this is probably a fair ratio when compared with a number of other countries. However, the quality of education would certainly improve if the ratio were lower. For example, in many provinces in Canada the ratio is 4:1 and a teacher aide can be employed if there are more than 4 students – if only all countries were so benevolent. Hong Kong may be the wealthiest 400 square miles in the world, bearing the name of a developed country: why does it lag behind in this area of human concern?

2. Think normal

A constant stumbling block is the often negative, pessimistic attitudes of teachers, parents, and members of the public which has detrimental effects on the lives of people with mental retardation. Unless active and systematic measures are taken to "educate" the public, to convince them that special education is worth more real investment, advances in education for students with mental retardation will continue to be very difficult. Ordinary people, parents, teachers and other professionals have to understand that it is "normal" to be retarded. When given a better quality of training, students with mental retardation, regardless of the level of retardation, will make significant progress, and this makes economic as well as social sense.

Some students with mild mental retardation, who are capable of holding open employment or receiving more advanced training in vocational training centres, are working in sheltered workshops which are designed mainly for people with moderate mental retardation. On the other hand, the majority of students with moderate mental retardation (those who are capable of holding jobs in sheltered workshops) "work" in Day Activity Centres which primarily cater for people with **severe** mental retardation. Thus, instead of making advances in the employment sector, people with mental retardation are often fixed in situations which are not commensurate with their potential abilities. A bottleneck is created, not by possible employees, but by a serious underestimate of the potential ability of this group and a lack of consideration for the humane principle of normalization. Crowding students with mental retardation into special schools will not make education **special**. It is high time we in Hong Kong made sensible plans for integration.

3. Quality Training

Do competent teachers have good methods and incompetent teachers bad ones? Or, do good methods and bad methods depend entirely on competent teachers and incompetent teachers? One thing is certain: **all** teachers need quality training, perhaps at first degree but certainly at post graduate degree level. Piece-meal staff development programmes or attending a conference or two are stimulating but inadequate in developing high quality teachers. The Warnock Committee (1978) recommended in their report that **all** teachers in training should have an element of special education, and for teachers of students with mental retardation it should be a major element of their training, supplemented by post-graduate courses. In more then ten years we have done no more than **agree** with the recommendation.

Conclusion

In special education, the issues of **what is taught** and **how appropriately** have been largely ignored. The question posed here is: "Are we teaching what the students need to learn?" Alice Metzner (cited in Blatt, 1970) wrote: **"There are only two things wrong with most special education for the mentally handicapped, it isn't special, and it isn't education."** At best there is a watered-down version of the standard curriculum provided for most children-but with less of the same for a population whose rate of learning is substantially slower. One of the best indicators of the effectiveness or relevance of school training is to ask simply: Where will they go after they leave school? The fact that most do not hold meaningful employment, or are forced to linger in school as "mature" students, or just pass time idling at home (Hong Kong Association for the Mentally Handicapped, 1987 and 1988; Hong Kong Council of Social Service and Rehabilitation Coordinating Committee, 1987) shows that there is something wrong in the teaching content and/or its process. Teachers accumulate a good deal of knowledge about how to teach, through experience over years (Tse, 1985). Perhaps it is time we asked: Are the teaching activities good for the people they teach? Or: are they teaching the tasks that their students need to learn? We may come up with better answers to these questions when we start to draw the links between assessment and training more closely and we begin to pay more attention to the issue of **relevance** in what we teach.

To decide on "what to teach" is neither simple nor easy. The curriculum (or priority teaching areas) must aim at teaching skills that are required in the environments in which students with mental retardation live now and in which they are likely to be living in the future. Perhaps the most appropriate action is to build an ecological inventory. Designing an ecological curriculum does not only depend upon the development of a set of instructional outcomes that the students are capable of mastering. Although it is essential

that the students can master the subject matter, they also need to acquire the skills. The desire for relevant or functional curricula for students with mental retardation can create a problem unless teachers recognize that the skills of personal, social, motor, recreational, vocational, and home management must be learned.

Recently some special educators have voiced the opinion that education in a special school setting should not include vocational training. May I just quote Webster's Ninth New Collegiate Dictionary which defines the word "educate" as "to train by formal instruction and supervised practice especially in a skill, trade, or profession". *Res ipsa loquitur*, or let the words speak for themselves. Increasingly more professionals are advocating the inclusion of vocational skills training in special schools and the use of employment as a graduation goal (Hasazi and Clark, 1988; Lane, 1985; Rusch, Chadsey-Rusch, and Lagomarcino, 1987; Wehman and Kregel, 1985). Given appropriate training, students with mental retardation, together with the required positive expectations from teachers, supported or even competitive employment should be reasonable and attainable targets (Crouch, Rusch, and Karlan, 1984; Lane, 1985; Rusch and Hughes, 1988; Rusch, Martin, and White, 1985). Critical appraisal is a necessity if we are going to eliminate the traditional (or cosmetic) basis for deciding what will be taught. The intent is not to remove the 3 R's or more academic elements from the curriculum altogether. On the contrary, all students with mental retardation should be given opportunities to learn academic skills that they will be able to use effectively.

Special education teachers and principals can no longer continue to ignore the implications of mental retardation as it relates to long-term vocational employment and community adjustment. Those of us who are concerned with the education of students with mental retardation must make the enormous value decision of what to teach based upon the characteristics and needs of the students as these are related to their future as adults.

References

Baine, D. (1986). *Testing and teaching handicapped children and youth in developing countries*. Paris: UNESCO.

Balthzar, E. (1976). *Balthazar Scales of Adaptive Behaviour*. Palo Alto, California: Consulting Psychologists Press.

Barker, R. (1968). *Ecological psychology*. Stanford, California: Stanford University Press.

Blatt, B. (1970). *Exodus from pandemonium: Human abuse and a reformation of public policy*. Boston, MA.: Allyn & Bacon, Inc.

Brooks, P. H., & Baumeister, A. A. (1977). A plea for consideration of ecological validity in the experimental psychology of mental retardation: guest editorial. *American Journal of Mental Deficiency*, *81*, 407–516.

Brown, L., Hamre-Nietupski, S., Pumpian, I., Certo, N., & Gruenewald, L. (1979). A strategy for developing chronological-age-appropriate and functional curricular

content for severely handicapped adolescents and young adults. *The Journal of Special Education*, *13*, 81–90.

Brown, L., Nietupski, J., & Hamre-Nietupski, S. (1976). The criterion of ultimate functioning and public school services for severely handicapped students. In M. Thomas (ed.), *Hey don't forget about me: Education's investment in the severely, profoundly and multiply handicapped*. Reston, VA: Council for Exceptional Children.

Crawford, N., & Tse, J. (Eds.). (1988). *Home-based training package for persons with mental handicap*. Hong Kong: Social Welfare Department, Hong Kong Government.

Crouch, K. P., Rusch, F. R., & Karlan, G. P. (1984). Competitive employment: Utilizing the correspondence training paradigm to enhance productivity. *Education and Training of the Mentally Retarded*, *19*, 268–275.

Dalton, A., Cibiri, S., Baker, J., Malik, H., & Wu, B. (1981). *Basic Life Skills Scale*. Surrey Place, Ontario: Ministry of Community and Social Services. (Available from Ministry of Community and Social Services, Surrey Place Centre, 2 Surrey Place, Toronto, Ontario, Canada. Single copy free of charge).

Dell, D., Felce, D., Flight, C., Jenkins, J., & Mansell, J. (1986). *The Bereweeke Skill-teaching System*. (Revised ed.) Windsor, Berks.: NFER-Nelson.

Gannon, P. (1986). Research with moderately, severely, profoundly retarded and autistic individuals (1975 to 1983): An evaluation of ecological validity. *Australia and New Zealand Journal of Developmental Disabilities*, *12*, 33–53.

Hammill, D. (1982). Assessing and training perceptual-motor skills. In D. Hammill and N. Bartel (Eds.), *Teaching children with learning and behaviour problems*. Boston: Allyn and Bacon.

Hasazi, S., & Clark, G. (1988). Vocational preparation for high school students labeled mentally retarded: Employment as a graduation goal. *Mental Retardation*, *26*, 343–349.

Holvoet, J., Guess, D., Mulligan, M., & Brown, F. (1980). The individualized curriculum sequencing model (II): A teaching strategy for severely handicapped students. *Journal of the Association for the Severely Handicapped*, *5*, 337–351.

Hong Kong Association for the Mentally Handicapped (1987). *Employment situation of graduates of special schools for mentally handicapped*. Hong Kong: Author.

Hong Kong Association for the Mentally Handicapped (1988). *Employment situation of graduates of special schools for mentally handicapped*. Hong Kong: Author.

Hong Kong Council of Social Service and Rehabilitation Development Co-ordinating Committee. (1987). *Survey on employment situation of the disabled*. Hong Kong: Authors.

Kwok, J., Shek, D., Tse, J., & Chan, S. (1989). *Hong Kong based Adaptive Behaviour Scale*. Hong Kong: Department of Applied Social Studies, City Polytechnic of Hong Kong.

Lane, D. (1985). After school-work. In D. Lane & B. Stratford (Eds.), *Current approaches to Down's syndrome*. Eastborne: Holt, Rinehart & Winston.

Liberty, K., Haring, N., & Martin, M. (1981). Teaching new skills to the severely handicapped. *Journal of the Association for the Severely Handicapped*, *6*, 5–13.

Mattison, M., & Rosenberg, R. (1985). The use of an ecological inventory to select education program objectives. In M. Brady and P. Gunter (Eds.), *Integrating moderately and severely handicapped learners: strategies that work*. Springfield, Ill.: C. C. Thomas.

Miller, W. M. (1964). *A Canticle for Leibowitz*. Philadelphia, PA: J. B. Lippincott Co.
Nihira, K., Foster, R., Shellhaas, M., & Leland, H. (1974). *AAMD Adaptive Behaviour Scale*. Washington, DC: AAMD.
Myer, P. and Hammill, D. (1976). *Methods for learning disorders*. (2nd ed.) New York: Wiley.
Perske, R. (1972). The dignity of risk. In W. Wolfensberger (ed.), *The principle of normalization in human services*. Toronto: National Institute on Mental Retardation.
Rusch, F. R., Chadsey-Rusch, J., & Lagomarcino, T. (1987). Preparing students for employment. In M. Snell (ed.), *Systematic instruction of persons with severe handicaps*. (3rd ed.) Columbus: Charles E. Merrill Publishing Co.
Rusch, F. R., & Hughes, C. (1988). Supported employment: promoting employee independence. *Mental Retardation*, *26*, 351–355.
Rusch, F. R., Martin, J. E., & White, D. M. (1985). Competitive employment: Teaching mentally retarded employees to maintain their work behaviour. *Education and Training of the Mentally Retarded*, *20*, 182–189.
Seguin, E. (1846). *Le traitement moral, l'hygiene et l'education des idiots*. Paris: Bailliere.
Shavelson, R., Webb, N., & Burstein, L. (1986). Measurement of teaching. In M. Wittrock (ed.), *Handbook of research on teaching*. (3rd ed.) N. Y.: MacMillan Publishing Co.
Snell, M., & Grigg, N. (1987). Instructional assessment and curriculum development. In M. Snell (ed.), *Systematic instruction of persons with severe handicaps*. (3rd ed.) Columbus: Charles E. Merrill Publishing Co.
Snell, M., & Zirpoli, T. (1987). Intervention strategies. In M. Snell (Ed.), *Systematic instruction of persons with severe handicaps*. (3rd ed.) Columbus: Charles E. Merrill Publishing Co.
Somerton-Fair, E., & Turner, K. (1979). *Pennsylvania training model individual assessment guide*. (Revised ed.) Pennsylvania: Department of Education. (Available from Bureau of Special Education, Pennsylvania Department of Education, 210 East Fulton Street, Butler, Pennsylvania, USA 16001. Single copy free of charge).
Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behaviour Scales*. Circle Pines, Minnesota: American Guidance Service.
Special Education Division, Education Department. (1984). *Curriculum guidelines for mildly mentally handicapped children*. Hong Kong: Author.
Special Education Division, Education Department. (1985). *Curriculum guidelines for moderately mentally handicapped children*. Hong Kong: Author.
Special Education Division, Education Department. (1983). *Temporary curriculum guidelines for severely mentally handicapped children*. Hong Kong: Author.
Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behaviour Analysis*, *10*, 349–367.
Stratford, B. (1989). *Down's Syndrome: past, present and future*. Harmondworth: Penguin Books.
Thomas, D. (1985). Philosophical considerations for the curriculum of mentally retarded children. *Early Child Development and Care*, *22*, 123–136.
Thorndike, R. L. (1973). *Stanford-Binet Intelligence Scale, Form L-M, 1972 norms tables*. Boston: Houghton Mifflin.
Tse, J. (1985). *Tips on teaching mentally retarded persons*. Hong Kong: Society of Homes for the Handicapped.

Warnock Report. DES (1978). *Special educational needs. Report of the Committee of inquiry into the education of handicapped children and young people*. London: HMSO.

Wehman, P., & Kregel, J. (1985). A supported work approach to competitive employment of individuals with moderate and severe handicaps. *The Journal of the Association for Persons with Severe Handicaps*, *10*, 3–11.