

Planning and
Task Performance in
a Second Language

Edited by
Rod Ellis

John Benjamins Publishing Company

Planning and Task Performance in a Second Language

Language Learning and Language Teaching

The *LL<* monograph series publishes monographs as well as edited volumes on applied and methodological issues in the field of language pedagogy. The focus of the series is on subjects such as classroom discourse and interaction; language diversity in educational settings; bilingual education; language testing and language assessment; teaching methods and teaching performance; learning trajectories in second language acquisition; and written language learning in educational settings.

Series editors

Nina Spada

Ontario Institute for Studies in Education, University of Toronto

Jan H. Hulstijn

Department of Second Language Acquisition, University of Amsterdam

Volume 11

Planning and Task Performance in a Second Language

Edited by Rod Ellis

Planning and Task Performance in a Second Language

Edited by

Rod Ellis

University of Auckland

John Benjamins Publishing Company
Amsterdam/Philadelphia



™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Planning and task performance in a second language / edited by Rod Ellis.
p. cm. (Language Learning and Language Teaching, ISSN 1569-9471
; v. 11)

Includes bibliographical references and index.

1. Second language acquisition. 2. Second language acquisition--
Ability testing. 3. Second language acquisition--Methodology. 4.
Language and languages--Study and teaching. 5. Language planning. 6.
Competence and performance (Linguistics) I. Ellis, Rod. II. Series.

P118.2.P59 2005

418'.0071--dc22

2004066032

ISBN 90 272 1961 3 (Eur.) / 1 58811 613 1 (US) (Hb; alk. paper)

ISBN 90 272 1962 1 (Eur.) / 1 58811 614 X (US) (Pb; alk. paper)

© 2005 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Preface	VII
Section I. Introduction	
CHAPTER 1	
Planning and task-based performance: Theory and research	3
<i>Rod Ellis</i>	
Section II. Task rehearsal	
CHAPTER 2	
Integrative planning through the use of task-repetition	37
<i>Martin Bygate and Virginia Samuda</i>	
Section III. Strategic planning	
CHAPTER 3	
What do learners plan? Learner-driven attention to form during pre-task planning	77
<i>Lourdes Ortega</i>	
CHAPTER 4	
The effects of focusing on meaning and form in strategic planning	111
<i>Jiraporn Sangarun</i>	
CHAPTER 5	
The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency	143
<i>Chieko Kawauchi</i>	

Section IV. Within-task planning	165
CHAPTER 6	
The effects of careful within-task planning on oral and written task performance	167
<i>Rod Ellis and Fangyuan Yuan</i>	
CHAPTER 7	
Strategic and on-line planning: The influence of surprise information and task time on second language performance	193
<i>Peter Skehan and Pauline Foster</i>	
Section V. Planning in language testing	
CHAPTER 8	
Planning for test performance: Does it make a difference?	219
<i>Catherine Elder and Noriko Iwashita</i>	
CHAPTER 9	
Strategic planning, task structure, and performance testing	239
<i>Parvaneh Tavakoli and Peter Skehan</i>	
Section VI. Conclusion	
CHAPTER 10	
Planning as discourse activity: A sociocognitive view	277
<i>Rob Batstone</i>	
References	297
Index	309

Preface

The purpose of this book is to bring together a series of articles on the nature of planning and its effects on task-based performance in laboratory, classroom and testing contexts. The idea for the book originated in a colloquium on this topic given at AILA Conference in Singapore in December 2002. Papers given by Bygate and Samuda, Elder and Iwashita, Ellis and Fanguan, and Sanguran were subsequently developed into chapters for this book. A number of other researchers (Batstone, Foster, Ortega, Kawauchi, Skehan, and Tavakoli) were later invited to submit chapters and did so.

Planning and its role in task-based performance are of both theoretical interest to second language acquisition (SLA) researchers and of practical significance to language teachers. In the case of SLA researchers, planning is important because it links in with the current interest in the role of attention in language learning. Whether learners plan strategically before they perform a task or engage in careful within-task planning, opportunities arise for them to attend to language as form, or as Ortega (Chapter 3) puts it 'form-in-meaning'. Thus, investigating planning serves as one way of studying what learners attend to and what effect it has on the way they use language. Further, it is also hypothesized that the kind of language use that learners engage in will influence the process of acquisition itself. Its significance for language teachers lies in the fact that planning is a relatively straightforward way of influencing the kind of language that learners produce. It serves, therefore, as an effective device for intervening indirectly in interlanguage development.

The predominant methodological paradigm in planning studies is experimental. That is, the task performance of learners who engage in planning of one kind or another is compared with a task performance where there is no opportunity for planning. This paradigm continues to be reflected in several of the studies reported in this book (e.g. the chapters by Kawauchi, Ellis and Yuan, and Skehan and Foster). It has proved very fruitful in demonstrating that planning does indeed affect the way in which learners perform a task. Nevertheless, this paradigm also has its limitations. It tells us nothing about what learners actually do when they are planning; it does not show us whether learners actually

do what they planned to do; and, more crucially, perhaps, it fails to recognize that planning and task-performance constitute social as well as cognitive activities.

Clearly, then, there is a case for broadening the paradigm to incorporate both a process element and to acknowledge the social nature of tasks. A number of the chapters in this book address planning as a process. Ortega extends her earlier research on tasks to examine the strategies that learners use when engaged in pre-task planning. Sanguran (Chapter 4) discusses how the instructions learners are given can influence the way in which they plan. Several authors report the results of post-task questionnaires designed to investigate how learners responded to the opportunities to plan. Skehan and Foster (Chapter 7) undertake a detailed analysis of what they call 'breakdown fluency' with a view to identifying process features of task performance that will provide evidence of on-line planning. All of these studies extend the research on planning in significant ways.

There is less evidence of any attention to the social aspect of planning and task-performance. The prevailing tenor of this book is psycholinguistic. In the concluding chapter, however, Batstone (Chapter 10) develops a convincing argument for a social perspective. He points out that learners can approach tasks in two different ways – as requiring economical and efficient communication or as providing opportunities for them to engage in learning activities. The idea that tasks always have a context and that this context will help to shape how learners plan for and perform them is further supported in the two chapters that address the role of task planning in a testing situation (by Elder and Iwashita [Chapter 8] and Tavakoli and Skehan [Chapter 9]). The very different results of these two studies are perhaps best explained in terms of the differences in the specific testing contexts.

It is to be hoped, then, that this book both reflects mainstream research into the role of planning in task-based performance and also extends it.

Rod Ellis
Auckland, April 2004

SECTION I

Introduction

The last decade has seen a growing body of research investigating various aspects of L2 learners' performance of tasks (see, for example, Bygate et al. (2001) and Ellis (2003)). This research has focused broadly on a variety of design features of tasks and implementation procedures and how these impact on such aspects of language use as comprehension, input processing, meaning negotiation and the fluency, complexity and accuracy of L2 production (Skehan 1996, 1998a). While task-based research has been able to identify a number of variables that impact on performance (e.g. whether contextual support is available, whether the information is shared or split, whether the outcome is closed or open, whether there is inherent structure to the task's content), the results have not always been consistent. This has led some researchers (e.g. Coughlan & Duff 1994) to argue that the 'activity' that results from a 'task' is necessarily co-constructed by the participants on each occasion, making it impossible to predict accurately or usefully how a task will be performed.

However, one implementation variable that has attracted considerable attention and that has been shown to produce relatively consistent effects on L2 production is task planning. A number of studies (e.g. Foster & Skehan 1996) have shown that when learners have the opportunity to plan a task before they perform it, the language they produce is more fluent and more complex than when no planning is possible. Other studies (e.g. Yuan & Ellis 2003) have shown that unpressured on-line planning also has predictable effects, albeit somewhat different from those arising from pre-task planning.

The choice of planning as the variable for investigation in this book is motivated both by its importance for current theorizing about L2 acquisition (in particular with regard to cognitive theories that view acquisition in terms of information processing) and its value to language teachers, for unlike many other constructs in SLA, 'planning' lends itself to pedagogical manipulation. The study of task planning, then, provides a suitable forum for establishing the interconnectedness of theory, research and pedagogy in SLA (Pica 1997).

This introductory chapter has a number of purposes. It seeks to provide a framework for the subsequent chapters by identifying and defining different

types of planning. It examines the theoretical backgrounds that have informed the study of planning in task-based performance. It reviews earlier research that has investigated the effects of the different types of planning. It examines a number of key methodological issues related to the study of the effects of planning on task performance.

CHAPTER 1

Planning and task-based performance

Theory and research

Rod Ellis

University of Auckland

Types of planning

All spoken and written language use, even that which appears effortless and automatic, involves planning. That is speakers and writers have to decide what to say/write and how to say/write it. Planning is essentially a problem solving activity; it involves deciding what linguistic devices need to be selected in order to affect the audience in the desired way. As Clark and Clark (1977) noted, planning takes place at a number of different levels, resulting in discourse plans, sentence plans and constituent plans, all of which have to be interwoven in the actual execution of a language act.

Principal types of task planning

Figure 1 distinguishes two principal types of task-based planning – *pre-task planning* and *within-task planning*. These are distinguished simply in terms of when the planning takes place – either before the task is performed or during its performance. Pre-task planning is further divided into *rehearsal* and *strategic planning*. Rehearsal entails providing learners with an opportunity to perform the task before the ‘main performance’. In other words, it involves task repetition with the first performance of the task viewed as a preparation for a subsequent performance. Strategic planning entails learners preparing to perform the task by considering the content they will need to encode and how to express this content. In pre-task planning, the learners have access to the actual task materials. It is this that distinguishes strategic planning from other types of pre-task activity (e.g. brainstorming content; studying a model performance of

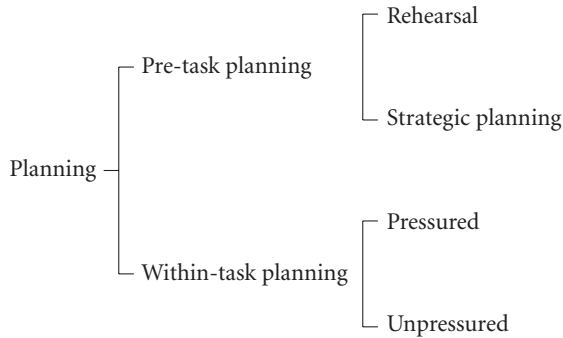


Figure 1. Types of task-based planning

the task; dictionary search). Within-task planning can be differentiated according to the extent to which the task performance is pressured or unpressured. This can be achieved most easily by manipulating the time made available to the learners for the on-line planning of what to say/write in a task performance. In an unpressured performance learners can engage in careful on-line planning resulting in what Ochs (1979) has called ‘planned language use’. In pressured performance learners will need to engage in rapid planning resulting in what Ochs calls ‘unplanned language use’ (although, of course, all language use involves some level of planning). Ochs documents a number of linguistic differences between the two types of discourse. For example, unplanned discourse tends to manifest non-standard forms acquired early whereas planned discourse contains more complex, target-like forms.

While pre-task planning and within-task planning constitute distinctive types of planning they should not be seen as mutually exclusive. As shown in Figure 2, it is possible to envisage four basic combinations of the two planning conditions. In condition 1, learners have no opportunity for either pre-task planning or unpressured within-task planning. Given that learners (especially with low proficiency) have a limited processing capacity and are likely to experience difficulty in accessing and encoding their linguistic knowledge, this can be considered the most demanding condition. In condition 2, learners are given the opportunity to pre-plan their performance (either by means of task rehearsal or strategic planning) but are not allowed to plan their utterances carefully on-line. In condition 3, the reverse occurs; learners are required to start performing the task straight away but are given as much time as they wish to take. Both of these conditions may ease the processing burden of the learner.

Planning conditions	Pre-task planning	Unpressured within-task planning
1	No	No
2	Yes	No
3	No	Yes
4	Yes	Yes

Figure 2. Planning conditions

Condition 4, where the learner has the opportunity for both pre-task planning and unpressured within-task planning, can be expected to create the conditions that help learners maximize their competence in performance.

Sub-categories of task planning

Both pre-task and within-task planning can be categorized further in ways not shown in Figure 1 but which are of potential theoretical and practical significance. For example, learners can be left to their own devices when planning a task (unguided planning) or they can be given specific advice about what and how to plan (guided planning). In this case, they can be directed to attend to linguistic form, to meaning or to form and meaning. Chapter 4 by Sangarun, for example, explores how directing learners to focus on some specific aspect of language in their strategic planning of tasks influences subsequent performance. Earlier studies (e.g. Hulstijn & Hulstijn 1984) have explored the effects of directing attention to form or meaning on within-task planning and performance. Another option relevant only to strategic planning concerns participatory structure, i.e. whether the planning is undertaken by the learners working individually, collaboratively in small groups, or with the teacher (see Foster & Skehan 1999). As Batstone discusses in the concluding chapter to this volume this can potentially affect the way a task is performed.

Clearly, which types and combinations of types of planning are of relevance must ultimately be decided empirically. That is, each type/option needs to be systematically examined to establish if it has any effect on the language produced in a task performance. As we will see when we examine the previous research on planning and task-based performance this has been one of the major goals of enquiry to date.

Theoretical background to the study of planning in task-based research

I will consider three theoretical frameworks that have informed the study of task planning in second language acquisition (SLA) research. These are (1) Tarone's (1983) account of stylistic variation, (2) models of speech production and writing, and (3) cognitive models of L2 performance and language learning. These theories explicitly or implicitly draw on three central constructs involved in psycholinguistic accounts of language processing – attention and noticing, a limited working memory capacity, and focus-on-form – so I will begin by briefly outlining each of these constructs, as they have been applied in SLA research.

L2 production as information processing: Some key constructs

Information processing models constitute the dominant approach to theorizing about language comprehension and production in cognitive psychology today. While the current models differ in some major ways (see Robinson 1995 for a review of these), they all share a number of features; they all seek to account for how information is stored and retrieved; they all view information processing as involving input, temporary storage of material attended to, long-term storage of (some of) this material and mechanisms for accessing this material from long-term memory. Lantolf (1996) has referred to this general approach as the 'computational model' as it is based on an analogy between the human mind and a computer.

There are a number of general principles that inform this model (Huitt 2003). One is the assumption of a limited capacity. That is, there are limits on the amount of information that human beings can process from input or for output. These limits cause bottlenecks in working memory and can lead to language users prioritizing one aspect of language over another. A second principle is that there is a control mechanism that language users will need to access when they are confronted with a new task for which they do not possess proceduralized linguistic knowledge. This control mechanism draws on explicit stored knowledge. As such, it uses up processing power and thus taxes working memory. A third principle is that human beings process information by means of both top-down processes that draw on encyclopedic knowledge of the world and on situational context and bottom-up processes that involve close attention to the linguistic signals in the input. These general principles underlie the three central constructs discussed below.

1. *Attention and noticing*

In a number of seminal articles in the 90s, Schmidt (1990, 1994) advanced the hypothesis that conscious attention, or what he called 'noticing', is essential for language learning. He states 'although unattended stimuli may have subtle but undeniable effects on humans (as in subliminal perception experiments), it is widely argued in psychology that *learning* without attention to what is to be learned is impossible' (Schmidt 1994:17). He goes on to argue that in the case of learning attention must necessarily be conscious as 'all demonstrations of detection without conscious registration . . . demonstrate only the processing of what is already known, not learning'. This is a view that has not gone unchallenged, however. In particular, Tomlin and Villa (1994) have proposed that three components of attention can be distinguished; alertness (a general readiness to deal with incoming stimuli), orientation (the aligning of the attentional mechanisms to some specific aspect of language) and detection (the actual process by which a specific feature of language is attended to focally). They claim that none of these components necessarily involves consciousness and that even detection can occur without any conscious registration of the stimuli attended to. More recently, Schmidt (2001) has been less dogmatic about whether (conscious) attention is required, writing 'the question of whether *all* learning from input requires attention to that input remains problematic, and conceptual issues and methodological problems have combined to make a definitive answer elusive' (p. 29). He continues to assert, however, that intentional, conscious attention is beneficial for learning as it can help learners process features of language that otherwise would not be noticed.

Much of the discussion of noticing (as conscious attention) in language learning has focussed on its role in input processing and, as such, might be seen as having little relevance to theorizing about how task planning aids acquisition. Task planning, whether of the pre-task or within-task type, may involve learners attending to the linguistic input provided in the task materials (e.g. in a text reformulation task), but in many tasks (e.g. those that involve a pictorial rather than verbal input) it clearly does not. Planning primarily entails learners accessing their own implicit and explicit knowledge of the L2 for use in production, as suggested by Ochs' (1979) account of planned language use. The question arises, then, as to whether noticing has any role to play in output-processing. Swain (1985b, 1995) claims that it does. According to the Output Hypothesis, production requires learners to process syntactically, which involves bottom-up rather than top-down processing and requires attention to form. Similarly, Robinson (2001b) suggests that output as well as input requires attention and that the degree of attention will depend on the complexity

of the task they asked to perform, with more complex tasks requiring more attention. Providing learners with the opportunity to plan a task, therefore, may aid performance. However, as we will see later, there is some disagreement as to how pre-task planning affects attention. One view is that it encourages greater attention to form during task performance, resulting in increased accuracy and complexity. An alternative view, promulgated by Robinson, is that pre-task planning simplifies the task and thus obviates the need to attend closely to form during performance but assists automatic access to stored language and so leads to greater fluency.

2. *Limited working memory capacity*

There are number of models of working memory (see Miyake & Shah 1999). One of the most commonly cited in the task planning literature is that of Baddeley (e.g. Baddeley & Hitch 1974; Baddeley & Logie 1999). This identifies three components of working (or short-term) memory; the central executive or supervisory attentional system, the phonological loop, and the visual spatial sketchpad. Two of these seem relevant to a role for task planning (i.e. not the visual spatial sketchpad).

The central executive system governs the relationship between working memory and long-term memory, allocating attention to specific long-term memory systems. This system is limited in capacity, and thus the extent to which language learners are able to attend to a specific system will depend on the extent to which other systems are automatized. For example, if learners use up available processing space in lexical searches the attention they can pay to grammar will be limited. Providing learners with the opportunity for pre-task planning or for unpressured within-task planning can ease the burden on working memory, allowing learners the opportunity to engage in controlled processing and to process multiple systems linearly.

The phonological loop is comprised of two sub-components – the phonological store, which affords a temporary representation of material drawn from the input or long term memory, and a mechanism that allows for articulatory rehearsal, which enables decaying material introduced into working memory to be sustained. Planning is likely to draw extensively on this component, allowing learners to maintain one set of material while drawing on another set to modify or refine it. For example, learners will be able to access linguistic material from their interlanguage grammars and maintain this in the phonological loop while they edit it through reference to their explicit knowledge of the L2. In other words, the phonological loop is likely to play a central role in monitor-

ing (discussed below). In short, planning is seen as a means of helping learners overcome the limitations in capacity of their working memory.

3. *Focus-on-form*

The term 'focus-on-form' has been variably used in the SLA literature. It helps to distinguish three related but different senses of the term, depending on whether the perspective is a pedagogic one, a discursual one or a psycholinguistic one. In the context of language pedagogy, focus-on-form refers to attempts to intervene in the process of acquisition by inducing learners to pay attention to linguistic form while they are primarily concerned with decoding or encoding message content. These attempts can be planned (i.e. a specific form is selected for attention) or incidental (i.e. specific forms are attended to as the need arises). In discursual terms, focus-on-form refers to the pre-emptive and reactive devices that interlocutors use to draw attention to form while learners are engaged in performing some task that gives priority to message conveyance. These devices can consist of 'queries' (i.e. questions about linguistic form) or various types of implicit and explicit corrective feedback (e.g. reformulations of learners' incorrect utterances, known as 'recasts', and metalinguistic explanation). In psycholinguistic terms, 'focus-on-form' refers to the mental processes involved in selective attention to linguistic form while attempting to communicate. 'Noticing', discussed above, serves as a cover term for these processes.

SLA researchers argue that L2 acquisition, especially in the case of adult learners, requires a focus-on-form. There are two rationales for this claim. The first relates back to the idea that learners have a limited working memory capacity and therefore experience difficulty in attending to meaning and form at the same time (see, for example, VanPatten 1990). Because it is 'natural' for learners to give priority to meaning, they may overlook certain linguistic features, especially those that are non-salient, redundant or do not contribute to meaning. As a result they need to be induced to attend to the formal aspects of the language. The second, more controversial claim is that interlanguage development can only take place if learners attend to form *while* they are engaged with meaning. As Doughty and Williams (1998) put it 'the fundamental assumption of FonF instruction is that meaning and use must already be evident to the learner at the time that attention is drawn to the linguistic apparatus needed to get the meaning across' (p. 4). They propose that there is a 'cognitive window for the provision of focus on form' of up to 40 seconds; that is, learners are able to hold material in working memory for this length of time during which they have the opportunity to attend to the form of what they have tem-

porally stored. Doughty (2001) suggests that 'roving attention' enables learners to pay attention to form without interruption of their original speech plan. The theoretical and empirical bases for these proposals are reviewed in detail in Doughty (2001).

Providing learners with the opportunity to plan a task performance constitutes a means of achieving a focus-on-form pedagogically. It mitigates the limitations of their working memory by allowing learners the 'cognitive window' needed to attend to form while they are primarily concerned with message conveyance. In other words, it creates a context in which learners have the opportunity to map form onto meaning by accessing linguistic knowledge that is not yet automatized.

Theoretical bases for task planning

The three constructs discussed above all figure to a greater or lesser extent in the theories of language use/acquisition that I will now consider. The three theories to be considered are presented chronologically, reflecting their origins in the history of task-based research. In each case I will outline the theory and then consider how it has been applied to task planning.

1. *Tarone's theory of stylistic variation*

Tarone's theory draws heavily on Labov's account of stylistic variation in native speakers. Labov (1970) argued that 'there are no single style speakers'; that is, individual speakers manifest variation in their use of language because they are able to draw on a variety of 'styles'. Further, he argued that 'these styles can be ranged along a single dimension according to the amount of attention that speakers pay to their speech' (i.e. focus on form). Depending on the situation, speakers vary in the extent to which they monitor their speech. Attention through monitoring is greatest in speech that reflects a careful style and least in the vernacular style found in everyday speech. Labov was able to show that what he called 'style shifting' was probabilistic but also systematic and therefore predictable. That is, speakers tended to use one variant in one style and another variant in another style to a greater or lesser extent depending on whether the social context encouraged them to pay attention to what they said.

Drawing on this theory of intra-speaker variability, Tarone (1983) proposed what she called the Capability Continuum for L2 learners. This consists of a continuum of styles, ranging from the 'careful' to the 'vernacular', which Tarone saw as comprising the learner's L2 knowledge. To explain how L2 development takes place, Tarone proposed two ways in which new forms can enter

interlanguage. In one way, forms originate in the learner's vernacular style and then spread to the more careful styles over time. In the other way, forms appear initially in the learners' most careful style, manifest only when the learner is paying close attention to speech production, and then spread to the less formal styles where they replace earlier, more primitive forms. Subsequent empirical work (e.g. Tarone 1985; Tarone & Parrish 1988) was directed at showing how the choice of forms was strongly influenced by the nature of the task learners were asked to perform. However, contrary to expectations, these studies did not always show that the more target-like forms occurred with greater frequency in tasks designed to elicit a careful style.

Viewing learners' L2 knowledge as a 'capability continuum', then, can explain how planning assists L2 production and acquisition. In the case of unpressured online planning, as in conditions 3 and 4 in Figure 2, learners will be able to attend to their speech and thus access their careful style. This will be reflected in greater accuracy (i.e. a more target-like performance). However, the provision of opportunity for careful on-line planning may not in itself promote acquisition. In this respect, pre-task planning followed by the pressured performance of a task (i.e. condition 2 in Figure 2) may be more effective. Pre-task planning allows learners to access their careful style but then requires them subsequently to perform the features they have accessed in real time where close attention to speech is not possible, thus encouraging the spread of these features from the careful to the vernacular style.

Nevertheless, the theory lacks explanatory power. First, it does not account for why some forms are more target-like in the learner's vernacular style. Second, the role of attention is not clearly specified. Third, the key notion of 'spread' is underdeveloped. The theory originated in a social account of language variation but planning is essentially a psycholinguistic construct. Current research on the role of planning has turned to theories that offer a fuller psycholinguistic account of L2 production.

2. Models of speech production and writing

By far the most influential theory where studies of task planning are concerned is Levelt's (1989) model of speech production. Many of the later chapters (i.e. Chapters 2, 4, 6, 7 and 9) draw on this model. The model was developed to account for the speech production of native speakers but de Bot (1992) has adapted it for bilingual speech production.

Levelt's (1989) model identifies three autonomous processing stages: (1) conceptualizing the message, (2) formulating the language representation, and (3) articulating the message.

The **conceptualizing** stage involves three sub-stages. First, the speaker decides upon the communicative goal. In the second substage (macro-planning) the speaker develops the communicative goal into a series of sub-goals and then identifies a speech act for each sub-goal that will achieve the intended effect. In the third sub-stage (micro-planning), the speaker retrieves the information needed to realize each of the subgoals and organizes it by determining ‘the information perspective of [an] utterance, its topic, its focus, and the way in which it would attract the addressee’s attention’ (Levelt 1989:5). The product of the micro planning is a preverbal message that is not linguistic in nature but contains, nonetheless, all information needed to convert the preverbal message into language. This preverbal message is then forwarded to the formulator.

Formulation involves establishing language representations of the preverbal messages by retrieving lexical items from the speaker’s mental lexicon. Each lexical item is comprised of two kinds of information: ‘lemma’ and ‘lexeme’. The lemma contains information about the meaning and syntax of each lexical item, while the lexeme contains information about its morphological and phonological properties. Thus, retrieving a lexical item serves to prompt the syntactic building procedure required for grammatical encoding. This results in a ‘surface structure’ (i.e., ‘an ordered string of lemmas grouped in phrases and subphrases of various kinds’ (Levelt 1989: 11)), which is then processed by the phonological encoder, resulting in a phonetic or articulatory plan (i.e., “an internal representation of how the planned utterance should be articulated” (Levelt 1989:12)). Levelt (1989) calls this ‘internal speech’.

Finally, this internal speech is transferred to the **articulator**. The articulator retrieves chunks of internal speech that are temporarily stored in an articulatory buffer and then “unfolds and executes [them] as a series of neuromuscular instructions” (p. 27). This leads, ultimately, to the production of overt speech.

These three stages are regulated by a self-monitoring process consisting of three subsystems. The first subsystem inspects whether the preverbal message matches the speaker’s original intention. It does this before the message is sent on to the formulator to be converted into internal speech. The second subsystem inspects the internal speech before it is articulated as overt speech. Finally, the third subsystem inspects the overt speech that has been generated.

Levelt (1989) also identified two characteristics of speech production which are relevant to task planning; (1) controlled and automatic processing and (2) incremental production. According to Levelt, some of the components of the speech production process (specifically, the conceptualizer and the monitor) operate under controlled processing, while other components (specifically, the formulator and the articulator) operate automatically in the

main. In addition, he proposed that speech production processes can take place in parallel.

De Bot (1992) considers the adaptations to Levelt's model needed to account for speaking in an L2. He suggests that in the case of the conceptualizer, macro-planning is not language specific but micro-planning is (i.e. the pre-verbal message specifies which language (or languages) are to be used to encode the message). De Bot argues that there are separate systems for the L1 and L2 as far as the processing components of the formulator are concerned, although the two systems are likely to be connected in at least some areas. In contrast, given the cross-linguistic influences evident in L2 pronunciation, he considers the existence of two separate systems for articulation 'very improbable' (p. 17). We might also note that whereas L1 speakers are able to carry out the processes involved in formulation and articulation (but not conceptualisation) without attention, L2 learners (especially those with limited L2 proficiency) are more likely to need to activate and execute their linguistic knowledge through controlled processing. Thus, they are likely to experience problems during the formulation and articulation stages, as these processes are demanding on working memory.

Levelt's model is explicitly designed to account for speech production. However, available theories of writing (e.g. Bereiter & Scardamalia 1987; Hayes & Flower 1980; Grabe 2001; Grabe & Kaplan 1996; Kellog 1996; Zimmerman 2000) posit a very similar set of processes to those proposed by Levelt. There is also general acceptance that these processes will be broadly similar in both L1 and L2 writing. Kellog's (1996) model, for example, distinguishes three basic systems involved in written text production. Each system has two principal components or processes. **Formulation** entails (1) 'planning', where the writer establishes goals for the writing, thinks up ideas related to these goals and organizes these to facilitate action, and (2) 'translating', where the writer selects the lexical units and syntactic frames needed to encode the ideas generated through planning and represents these linguistic units phonologically and graphologically in readiness for execution. **Execution** requires (3) 'programming', where the output from translation is converted into production schema for the appropriate motor system involved (e.g. handwriting or typing) and (4) 'executing', the actual production of sentences. **Monitoring** consists of (5) 'reading', where the writer reads his or her own text ('a necessary but not sufficient condition for writing well', p. 61) and (6) 'editing', which can occur both before and after execution of a sentence and can involve attending to micro aspects of the text such as linguistic errors and/or macro aspects such as paragraph and text organization. The extent to which a writer is able to engage in monitoring will depend

in part on whether the writer has the time to adopt a 'polished draft strategy' or is engaged in pressured text production. Kellogg, like Levelt for speaking, emphasises that writers simultaneously activate formulation, execution, and monitoring processes, although the extent to which this is achievable depends on working memory.

Kellogg also suggests how the different components of the model relate to working memory. He argues that the central executive, a multi-purpose system responsible for problem-solving (see above), mental calculation and reasoning, is involved in all the sub-processes with the exception of executing, which, he argues, is usually accomplished without the need for controlled processing. It should be noted, however, that this assumes an adult, native-like automaticity in handwriting or typing, which may be lacking in some L2 learners, especially those whose first language (L1) employs a different script. It is possible, therefore, that the central executive may be called upon by some L2 writers during execution. Kellogg suggests that the visuo-spatial sketchpad, which stores and processes visual and spatial information in working memory, is only involved in planning. Finally, he proposes that the phonological loop, which stores and processes auditory and verbal information, is required for both translating and reading. The key feature of Kellogg's model is that the central executive has limited capacity, with the result that a writer may have to make decisions about which writing process to prioritise when under pressure to produce text rapidly. This is reflected in a trade off of attention directed at the different processes. Formulation demands are seen as critical, taking priority over execution and monitoring.

These models provide a basis for considering what components of language production (spoken or written) learners focus on while planning and also for examining what effects planning strategies have on actual production. Rehearsal, for example, may provide an opportunity for learners to attend to all three components in Levelt's model – conceptualisation, formulation and articulation – so it would seem reasonable to assume that this type of pre-task planning will lead to all-round improvements when the task is repeated, as found by Bygate (1996). Strategic planning can be considered likely to assist conceptualisation in particular and thus contribute to greater message complexity and also to enhanced fluency, as found by Wendel (1997). Unpressured within-task planning may prove beneficial to formulation and also afford time for the controlled processing required for monitoring. As a result, accuracy might increase. In other words different types of planning can be predicted to ease the pressure on the learner's limited working memory in different ways,

variably affecting the competition and trade-offs evident in different aspects of language production, as claimed by Skehan and Foster (1997).

The main advantage of these models of language production, then, is that they offer a detailed description of what is involved in speaking and writing and thereby afford relatively precise hypotheses about the effects that planning will have on task performance. In one respect, however, they are more limited than Tarone's theory of stylistic variation; they do not account for how linguistic change takes place, for, as Levelt (1989) has pointed out, they constitute steady-state models. Thus, while the models can explain the relationship between planning and language use they do not address how language use contributes to language acquisition.

3. *Cognitive models of task-based performance and learning*

Skehan's (1998b) 'cognitive approach' is based on a distinction between an exemplar-based system and a rule-based system. The former is lexical in nature and includes both discrete lexical items and ready-made formulaic chunks of language. The linguistic knowledge contained in this system can be easily and quickly accessed and thus is ideally suited for occasions calling for fluent language performance. The rule-based system consists of abstract representations of the underlying patterns of the language. These require more processing and thus are best suited for more controlled, less fluent language performance. They are needed when learners have to creatively construct utterances to express meaning precisely or in sociolinguistically appropriate ways.

Skehan also distinguishes three aspects of production; (1) fluency (i.e. the capacity of the learner to mobilize his/her system to communicate in real time, (2) accuracy (i.e. the ability of the learner to perform in accordance with target language norms) and (3) complexity (i.e. the utilization of interlanguage structures that are 'cutting edge', elaborate and structured). He suggests that language users vary in the extent to which they emphasize fluency, accuracy or complexity, with some tasks predisposing them to focus on fluency, others on accuracy and yet others on complexity. These different aspects of production draw on different systems of language. Fluency requires learners to draw on their memory-based system, accessing and deploying ready-made chunks of language, and, when problems arise, using communication strategies to get by. In contrast, accuracy and, in particular, complexity are achieved by learners drawing on their rule-based system and thus require syntactic processing. Complexity is distinguished from accuracy in that it is related to the 'restructuring' that arises as a result of the need to take risks whereas accuracy reflects the learner's attempt to control existing resources and to avoid errors.

Whereas Skehan's research assumes that learners possess a limited processing capacity such that trade-offs between fluency, accuracy and complexity (especially these last two) are likely to occur, Robinson's (2001c) research is premised on a multiple-resources view of processing – that is, that learners, like native speakers, have the capacity to attend to more than one aspect of language at the same time. According to this view, structural complexity and functional complexity are not in competition, as Skehan claims, but are closely connected such that increasing the cognitive complexity of a task is hypothesized to lead to greater linguistic complexity and accuracy as a result of increased output modification and input incorporation.

In Robinson's theory, task complexity is determined by two sets of features, 'resource directing' (e.g. whether or not the task requires reasoning) and 'resource depleting' (e.g. whether or not there is opportunity for strategic planning). These two factors 'interact and affect task production in measurable ways' (p. 31). Optimal attention to form arises when the task is resource directing and not resource depleting, as would be in the case when a task requires reasoning and there is no opportunity for strategic planning. Robinson argues that such a task is likely to enhance complexity and accuracy at the expense of fluency. In contrast a simple task that has no reasoning demands and allows opportunity for strategic planning is likely to promote fluency but not accuracy or complexity.

It is clear, then, that Skehan's and Robinson's models afford contradictory predictions as to the effects of planning on language performance. According to Skehan's model, strategic planning provides an opportunity for learners to access their rule-based system and thus makes them less reliant on their exemplar-based system. It may also assist them in taking the risks needed to access 'cutting edge' interlanguage features rather than relying, conservatively, on more fully acquired features. Thus, it is predicted to enhance linguistic complexity to the detriment of accuracy. In contrast, in Robinson's model, strategic planning is seen as a resource-depleting factor that works hand in hand with negative resource-directing factors to determine the overall complexity of the task and the extent to which learners attend to form when they perform the task, resulting potentially in increased fluency but decreased accuracy and complexity. However, as Robinson (2001b) admits the majority of studies of strategic planning have not supported his claim as they indicate a positive effect on complexity and, sometimes, on accuracy (see the section following). Neither Skehan nor Robinson consider the effects of unpressured on-line planning but presumably this can be hypothesized to work in similar ways to strategic planning (but see Skehan and Foster's chapter in this book).

Type of planning	Message content (Conceptualisation)	Formulation	Monitoring
1. Pre-task planning	Yes	Yes	No
2. Unpressured on-line planning	No	Yes	Yes

Figure 3. Planning and task performance

What insights do these various theories provide about how planning (1) affects task performance (spoken or written production) and (2) L2 acquisition? As shown in Figure 3, planning can impact on both the content learners communicate when performing a task and on their choice of language. In the case of the latter, planning is seen as important because of the role it can play in helping learners to access their L2 knowledge through controlled processing and, according to Skehan's theory, in promoting selective attention to form and monitoring. However, in accordance with the above discussion, the two principal types of planning – pre-task planning and unpressured on-line planning can be seen as impacting somewhat differently on these dimensions of performance. Thus, whereas pre-task planning contributes to the conceptualization of message content while also assisting controlled processing and selective attention to form, unpressured on-line planning has little impact on message content but facilitates language choice in formulation by allowing for controlled processing and selective attention to form and also monitoring.

While the theories are informative about how planning influences the performance of tasks, they are less convincing about how it contributes to acquisition. Extrapolating from performance to acquisition requires acceptance of a number of underlying assumptions:

1. Interlanguage development occurs while learners are primarily focused on message conveyance (i.e. performing tasks).
2. Interlanguage development is facilitated by selective attention to form.
3. Because learners have a limited working memory capacity, attention to form requires opportunity for controlled processing.
4. As a result of the opportunity for the selective attention made possible by controlled processing, learners are able to access more 'advanced' linguistic forms during the formulation stage of production and to achieve greater accuracy through monitoring than is possible in automatic processing.
5. One aspect of language use that fosters acquisition is the production of language that is complex and accurate (cf. Swain's Output Hypothesis).

These assumptions appear inherently reasonable, but, as we will see when I review the extant research on task planning, there is as yet very little empirical evidence in support of them. In particular, there is a notable lack of support for assumption 5, which is fundamental to the claim that planned language use assists acquisition.

Previous research on task planning

In line with the preceding typology of planning types, I will review the previous research on task planning by considering studies that have investigated pre-task planning and unpressured on-line planning. Studies of task-planning in a testing context will be considered separately.

Pre-task planning

1. *Rehearsal*

The research on rehearsal suggests that it has a beneficial effect on learners' subsequent performance of the same task but that there is no transference of the rehearsal effect to a different task, even when this is the same type as the original task. Bygate (1996) compared one learner's retelling of a Tom and Jerry cartoon on two separate occasions, three days apart. He found that rehearsal enhanced complexity, with the learner using more lexical verbs (as opposed to copula), more regular past tense forms (as opposed to irregular), a wider range of vocabulary and cohesive devices (e.g. words like 'then', 'so' and 'because'), and fewer inappropriate lexical collocations on the second occasion. There were also more self-correcting repetitions on the second telling of the story. Bygate (2001) reports a larger study that sought to investigate the effects of practicing specific types of task (involving narrative and interview) on both a second performance of the same task and on performance of a new task of the same type. The study showed that the second performance manifested greater fluency and complexity and also that the opportunity to practice that particular type of task helped. However, the practice did not appear to assist performance of a new task of the same type. In other words, disappointingly, there was no transfer of practice effect. Gass et al. (1999) report very similar findings in a study that compared learners' use of L2 Spanish in tasks with the same and different contents. In this study an effect for task repetition on ratings of overall proficiency, accuracy in the use of 'estar' (to a lesser extent) and

lexical complexity (type-token ratio) was found. However, again there was no transfer of these effects to a new task.

Lynch and McLean (2000; 2001) made use of a unique task that involved rehearsal. In the context of an English for specific purposes course designed to prepare members of the medical profession to give presentations in English, they designed a 'poster carousel' task. This required students to read an academic article and prepare a poster presentation based on it. Students then stood by their posters while other members of the group visited and asked questions. Altogether, each 'host' had six 'visitors'. Given that visitors tended to ask the same questions, there was substantial opportunity for retrieval. Lynch and Mclean document how recycling output resulted in both greater accuracy and fluency. However, they noted that different learners appeared to benefit in different ways with level of proficiency the key factor. Thus, whereas a learner with low proficiency appeared to benefit most in terms of accuracy and pronunciation, a learner with higher proficiency used the opportunity for retrieval to improve the clarity and economy of her explanations of a complex idea. Lynch and McLean also report considerable variation in the learners' awareness of the changes they were making in their production.

Task rehearsal, then, seems to have beneficial effects on learner performance. As Bygate (1999) suggests, learners are likely to initially focus on message content and subsequently, once message content and the basic language needed to encode it has been established, to switch their attention to the selection and monitoring of appropriate language. Bygate suggests that rehearsal may afford learners the extra processing space they need 'to integrate the competing demands of fluency, accuracy and complexity'. Bygate and Samuda, in Chapter 2, provide further evidence of this. However, it may not be inevitable that learners switch attention from content to form on the second performance. Nemeth and Kormos (2001) found that repeating an argumentative task influenced the number of supports the participants provided for their claims but had no effect on the frequency with which lexical expressions of argumentation were used. Also, before any strong claims can be made for rehearsal it will be necessary to show that the gains evident from repeating a task transfer to the performance of new, similar tasks.

2. *Strategic planning*

The role of strategic planning has attracted considerable attention from researchers. An effect on all three dimensions of production – fluency, accuracy and complexity – has been found. Each dimension will be considered sepa-

rately. First, though, I will consider research that has investigated what learners do when they plan strategically.

To date, only two studies have investigated what learners actually do when they are given the opportunity to plan. Wendel (1997) interviewed his learners immediately on completion of the tasks. They varied somewhat in what they reported doing during the planning time but all of them said they had focussed on sequencing the narrative events in chronological order. Only 3 reported attending to grammar but even these admitted it did not help them much when it came to telling the stories. As one learner put it: 'I feel like I'm pushing to tell you what's going on in the film. I focus on story, not grammar'. Wendel concluded that it is not useful for learners to try to plan the details of grammatical usage off-line. Ortega (1999) used retrospective interviews to investigate what learners did while they performed a narrative task. She found that they adopted an identifiable approach in their planning (e.g. they worked on the main ideas and organization first and then on the details), they attended to both content and linguistic form, and they made a conscious effort to plan at the utterance level. Ortega also reports that the learners varied considerably in the emphasis they gave to form and content, a point that she elaborates on further in Chapter 3. These two studies suggest that, when planning strategically learners tend to prioritize content. However, Ortega's study suggests that, not surprisingly, they do also attend to form.

Several studies indicate that strategic planning helps to enhance **fluency**. Studies by Foster (1996), Foster and Skehan (1996), Skehan and Foster (1997), Wendel (1997), Mehnert (1998), Ortega (1999) and Yuan and Ellis (1993) all report that giving learners the opportunity to plan results in greater fluency (i.e. a faster speaking rate and fewer dysfluencies). Foster (1996) and Foster and Skehan (1996) report that planners paused less frequently and spent less time in total silence than non-planners in all three tasks they investigated. However, the effect on fluency was stronger on the more difficult narrative and decision-making tasks than on the easier personal task. Skehan and Foster (1997), using similar tasks, replicated the result for total pauses. Wendel (1997) found that the planners in his study produced more syllables per minute and showed a lower mean length of pause in two narrative tasks. Ortega (1999) found a faster speech rate in learners of L2 Spanish on a story-telling task when they had an opportunity to plan strategically. Yuan and Ellis (2003) also report a clear effect for strategic planning on fluency. Foster (2001) found that planning resulted in learners producing a greater amount of speech whereas it led to native speakers producing less. Interestingly, Foster reports that the percentage of learner talk accomplished by means of lexicalised sequences did not change from the un-

planned to planned condition (i.e. it remained steady at about 17%) whereas that of the native speakers did change (from 32% in the unplanned to 25% in the planned). Her study suggests that planning opportunities may be used differently by learners and native speakers when the former lack an extensive store of lexicalized chunks and thus are forced to rely more on rule-based procedures in both planned and unplanned talk. Planning enables learners to access their rule-based procedures more speedily but not, so it would seem, to alter the balance of their use of formulaic and rule-based resources.

A question of obvious interest is what effect the amount of time allocated for planning has on fluency. A reasonable assumption is that the length of planning time is positively correlated with the degree of fluency. Mehnert (1998) set out to investigate this, allocating different groups of learners 0 minute, 1 minute, 5 minutes and 10 minutes of planning time. In general, she found that fluency did indeed improve in relation to the length of planning time. However, the main effect for fluency was evident between the non-planners and the planners; the differences among the three planning groups were mostly non-significant. Thus, providing learners with longer planning time did not have a major effect on the fluency of their speech.

In most of these studies, learners were simply given the task materials and told to plan what they wanted to say. However, a number of studies examined the effects of different kinds of strategic planning. Foster and Skehan (1996) investigated the effects of more guided planning. They compared the effects of 'undetailed' and 'detailed' planning, where the learners were given metacognitive advice about how to attend to syntax, lexis, content, and organization. The results showed that, in line with the overall effect of planning on fluency, for the narrative task the guided planners were notably more fluent than the unguided planners, but that there was no marked difference for the personal and decision-making tasks. This study suggests that the type of planning interacts with the type of task to influence fluency. Foster and Skehan (1999), however, found that asking learners to focus on form or meaning had no differential effect on fluency. Much may depend on the precise instructions given to the learners, as Sanguran (see Chapter 4) suggests. The study she conducted did find that focussing on form, meaning or form/meaning combined had an effect on fluency. Skehan and Foster also investigated the source of planning, comparing the effects of (1) teacher-led planning, (2) individual learner planning and (3) group-based planning on task performance. Where fluency was concerned, (2) proved most effective. However, as Batstone points out in Chapter 10, the ineffectiveness of the group-based planning may reflect the way in which the groups were constituted.

In contrast to fluency, the effects of strategic planning on accuracy are quite mixed. A number of studies reported that strategic planning led to increased accuracy. In Ellis (1987), I found that planning that provided opportunities for both strategic and on-line planning resulted in more accurate use of the regular past tense. Mehnert (1998) reported a significant difference in the accuracy of 1-minute planners over non-planners. However, the 5-minute and 10-minute planners performed at the same overall level of accuracy as the 1-minute planners. Other studies found no effect (e.g Crookes 1989; Wendel 1997). Yuan and Ellis (2003), using a general measure of accuracy, also found that strategic planning had no effect, a result that contrasted with that which they reported for unpressured on-line planning (see below). A number of studies found that strategic planning assisted accuracy only on some structures, some tasks and in some conditions. Ortega (1999) reported mixed findings – planning led to greater accuracy in the case of Spanish noun-modifier agreement but not in the case of articles. Foster and Skehan (1996) reported that both undetailed and detailed planners produced fewer errors than the non-planners on a decision-making task, that only the undetailed planners were more accurate than the non-planners on a personal task, while no effect for planning on accuracy was evident on a narrative task. Skehan and Foster (1997) found that planning (undetailed) led to greater accuracy on the personal and narrative tasks but not on the decision-making task. Foster and Skehan's (1999) study of the effects of source of planning found that accuracy was greatest when the planning was teacher-led. However, rather surprisingly, directing learners' attention to form as opposed to content during planning had no effect on accuracy.

It would appear from these results that whether strategic planning has any effect on accuracy may vary depending on a variety of factors, including the extent to which particular learners are oriented towards accuracy, the learners' level of proficiency, the type of task, the length of planning time available, and the particular grammatical feature. Also, with the exception of Yuan and Ellis (2003), these studies made no attempt to control for on-line planning. Thus, it is possible that the different results reflect whether learners were able to or chose to engage in monitoring while they performed the task. In terms of the Levelt model, strategic planning can be expected to aid conceptualisation but the impact of this may depend on the readiness of learners to shift attention to formulation when performing the task. If they do this, then strategic planning may lead to greater accuracy but if they do not do this no effect will be evident. Thus, the effect of strategic planning on accuracy may be linked to the kind of on-line planning that occurs subsequently during task performance. Clearly, though, more research is needed to identify how planning interacts with task

design variables, implementational procedures and individual learner factors. The variable impact of pre-task planning (rehearsal) on accuracy as a result of the learner's orientation during performance is explored by Bygate and Samuda in Chapter 2.

The results are clearer for **complexity**. As for fluency, strategic planning has a definite, positive effect; planners produce more complex language than non-planners. Crookes (1989) reports that 10 minutes of planning time led to learners producing more complex sentences and a wider range of lexis. Foster and Skehan (1996) found that detailed planners used significantly more subordination than undetailed planners who, in turn, produced significantly more subordination than the non-planners. This was broadly true for all three tasks. Skehan and Foster (1997), however, found that the planners' production was more complex on only two of the tasks. On the narrative task, where planning led to greater accuracy, no effect for complexity was evident, suggesting a trade-off between these two aspects of production. Wendel (1997) found that his planners used more complex grammatical structures but not more lexically rich language. Mehnert (1998) also found a positive effect but only for the 10-minute planners - the 1-minute and 5-minute planners performed at the same level as the non-planners. Ortega (1999) reports that mean number of words per utterance (a complexity measure) was significantly higher in the planning condition. Yuan and Ellis (2003) also found that strategic planning had a positive effect on complexity. With regard to the source of planning, Foster and Skehan (1999) found that individual learner planning worked best for complexity as it did for fluency. Again, in this study, whether the learners focused their planning on form or content had no differential effect on complexity.

These studies indicate that giving learners the opportunity to plan can increase the complexity of their production. They also suggest that this effect can be enhanced if (1) learners have a reasonable length of time to plan, say 10 minutes, (2) they are given guidance in how and what to plan and (3) they plan individually rather than in groups. It should be noted, however, that the measures of complexity used in these studies did not distinguish between propositional complexity (i.e. the content of the learners' messages) and formal complexity (i.e. the actual language used). Here too further research is needed.

What general conclusions are possible from these studies? The first is that strategic planning has a stronger effect on fluency and complexity than accuracy. This suggests that when learners plan strategically they give more attention to drawing up a conceptual plan of what they want to say rather than to formulating detailed linguistic plans. Even when asked to engage in form-focussed planning they may not do so, preferring to use the time given them

to sequence ideas and to work out the semantic linkages among propositions. Alternatively, it is possible that even when learners do attend to form when planning, they find it difficult to carry over the forms they have planned into the performance of the task, as suggested by Bygate and Samuda in Chapter 2. The second conclusion is that trade-off effects are evident. When learners plan they have to choose what aspect of production to focus on; focussing on fluency and complexity is at the expense of accuracy and vice-versa. Finally, there is some evidence to suggest that strategic planning has a greater effect on production in general when the task is cognitively demanding. If a task is easy learners are able to perform it fluently using accurate and complex language without the need for planning.

Unpressured on-line planning

Giving learners time to plan on-line and to monitor their output appears to have a clear impact on accuracy. Hulstijn and Hulstijn (1984) asked learners of L2 Dutch to perform short oral narratives under four conditions involving combinations of two variables; time (i.e. the learners were told to speak as quickly as they could or to take as much time as they wanted) and focal attention (i.e. learners were instructed to focus on form or on meaning). They found that time pressure by itself did not affect the accuracy of word order but that in combination with a focus on form it had a profound effect. This study, then, suggests that when learners use the time at their disposal to attend to formulation and to monitor the use of their grammatical resources their production becomes more accurate. However, if they use the time to plan content no effect on accuracy is observed.

In Ellis (1987), I compared learners' performance on written and oral narrative tasks based on pictures. In the case of the written task (task 1) the learners were given as much time as they wanted to write the narrative. In the first oral task (task 2) they were asked to retell the same narrative but without recourse to their written versions. In the second oral task (task 3) they were given a different set of pictures and instructed to tell the story with minimal opportunity for prior-planning. Figure 4 summarizes the kinds of planning opportunities afforded by these three tasks. I found that the learners' use of the regular past tense forms (but not the irregular past tense or copula past tense forms) was most accurate in task 1 and least accurate in task 3, with task 2 intermediate. The difference between task 1 and 2 can be explained in terms of on-line planning; accuracy was greater when there was no time pressure. However, as

Task	On-line planning/monitoring	Strategic planning
1	Yes	Yes (Probably)
2	No	Yes
3	No	No

Figure 4. Types of planning opportunities in Ellis (1987)

Crookes (1989) and others have pointed out, tasks 1 and 2 also differed with regard to medium.

Building on Ellis' study, Yuan and Ellis (2003) set out to compare the effects of pre-task and on-line planning on learner performance of a narrative task in a more systematic way. In the pre-task planning condition learners were given 10 minutes to prepare the task and then performed it under time pressure. In the on-line planning condition, the learners were given no chance to prepare but were allowed to perform the task in their own time. There was also a control group that had no preparation time and was required to perform the task under time pressure. The results indicated that opportunities for unpressured on-line planning assisted both accuracy and complexity but, as might be expected, inhibited fluency.

These three studies suggest that the time learners are given for on-line planning improves the accuracy of their production. However, the effects may only be evident when learners are drawing on their rule-based system. In both Hulstijn and Hulstijn (1984) and Ellis (1987) the effects of time pressure were only evident on grammatical structures that are clearly rule-based (i.e. Dutch word order rules and English regular past tense); they were not evident in structures that are more lexical in nature (i.e. irregular and copula past tense forms).

Planning in a language testing context

The study of the effects of planning on the performance of tasks in a testing situation is of considerable importance given that testers in general are concerned to elicit the 'best performance' from a testee (see McNamara 1996). If planning time can affect aspects of a test-taker's performance then arguably it ought to be considered when designing the test.

Three research studies have investigated the effects of pre-task planning in a testing situation. Wigglesworth (1997) examined the performances of 107 adult ESL learners performing five tasks that were part of the Australian Assessment of Communicative Skills (Access) test. The candidates performed the tasks in a planned and unplanned condition. The performances were rated

by two trained raters using an analytic rating scale to measure fluency, grammar (or in one task vocabulary) and intelligibility. The performances of 28 candidates, who were divided into high and low proficiency groups, were transcribed and analyzed using measures of complexity, fluency and accuracy. Wigglesworth reported no significant differences in the rating scores for the planned and unplanned conditions but significant differences in the analytic discourse measures for complexity, fluency and accuracy, especially in the high proficiency candidates and especially in tasks with a high cognitive load. She concludes that at least for some learners and in some tasks planning time can help to improve the performance of test-takers but that this effect is not evident in external ratings.

In a second study, Wigglesworth (2001) sought to further investigate one of the findings of the previous study, namely that the effects of planning time were not evident in the scores obtained from raters. The study examined the effect of a number of test task variables, one of which was planning, on adult ESL learners' performance on five tasks that were routinely used to evaluate achievement in the Australian Adult Migrant Education Program. In this study an effect for planning was found on the test-takers' ratings but the effect was not as great as might have been expected. Planning proved to have a detrimental effect on tasks that were familiar to the candidates and on both structured and unstructured tasks. Wigglesworth notes that these results are inconsistent with the findings of task planning research in non-testing situations and suggests that this may reflect the fact her study used external ratings rather than discourse analytic measures. However, Iwashita, Elder and Mcnamara (2001) used both analytic discourse measures and ratings to examine the effects of three minutes of planning time on the task performance of 201 ESL students and failed to find evidence of any effects on either the discourse measures or the rating scores. Elder and Iwashita reproduce this finding in Chapter 8 and examine a number of possible explanations.

It is possible, then, that the testing context constrains the beneficial effects of planning. This suggests, more generally, that the 'psychological context' of a task constitutes an important dimension that needs to be taken into account in planning studies (see Batstone's discussion of this possibility in Chapter 10). The main conclusion to be drawn from these studies, however, is that there is a need for further research into the effects of planning in a test situation. It seems clear, however, that whatever effect planning time has on task performance it may not be reliably measured by an external rating. This is problematic where assessment is concerned, as it is not practical to calculate discourse analytic measures in testing situations.

Final comments

This review of the research suggests that the effects of planning in a testing context may be somewhat different from those reported for laboratory or classroom contexts. One reason may be that learners feel pressured in a testing context with the result that their on-line planning is hurried. To date no studies have examined whether there are any differences in on-line planning in testing and non-testing contexts. This is a fairly obvious next step.

The results of the research certainly suggest that pre-task and unpressured on-line planning may be somewhat different. Whereas opportunities for on-line planning result in more accurate and complex language use, probably because learners have the chance to monitor linguistic form, opportunities for pre-task planning generally favour fluency and complexity, possibly because it leads to an emphasis on conceptualizing what has to be communicated rather than how to say it.

As I noted in the concluding comments to the previous section, researchers have focussed their attention on investigating how different types of planning (in combination with different types of tasks) impact on learner production. They have not attempted to show how or even whether the planning of tasks assists language acquisition. Thus any claims regarding planning and acquisition can only be theoretically based. Clearly, the absence of empirical support for the key assumptions listed at the end of the previous section constitutes a major lacuna in the research to date.

Methodological issues

The task planning research to date raises a number of methodological issues. Perhaps the key one concerns how acquisition as opposed to language production can be investigated. Other issues are how to ensure that learners carry out the type of planning specified in the research design and how to measure learners' actual production when they perform the task. These issues will be considered below.

Investigating the effects of planning on acquisition

The term 'acquisition' assumes that there is some change in the learner's L2 knowledge representation. Evidence for change can be found in (1) the learner's use of some previously unused linguistic forms, (2) an increase in

the accuracy of some linguistic forms that the learner can already use, (3) the use of some previously used linguistic forms to perform some new linguistic functions or in new linguistic contexts and (4) an increase in fluency (i.e. in the speed with which linguistic forms are used in communication).

The usual method for obtaining these kinds of evidence of change is the standard experimental design involving an experimental group that completes a pre-test, a treatment and post-tests (immediate and delayed) and a control group which receives the tests without the treatment. In the case of task planning research, the treatment consists of the opportunity to plan and perform a task. Such a design, as we have already seen is rarely employed. To the best of my knowledge, the only studies that have made use of such a design are Bygate's (2001) and Gass et al.'s (1999) studies of task rehearsal. Bygate's study asked learners in the experimental groups to perform two tasks prior to the treatment (which in turn consisted of three opportunities to repeat tasks similar to one of the pre-treatment tasks) and the same two tasks following the treatment together with two new tasks of the same type. In this way, Bygate was able to assess to what extent the treatment resulted in changes in the way the learners (1) performed the same task they had completed before the treatment and (2) a similar task to the pre-treatment task. Such a design is promising as it does allow the researcher to pinpoint changes that occur as a result of the treatment. It contrasts with the standard design used in task planning research (see, for example, Foster & Skehan 1996; Yuan & Ellis 2003), which typically involves an experimental and control group performing the same task under different planning conditions (e.g. strategic planning as opposed to no planning). Such a design cannot address acquisition.

There is, however, a major limitation to the kind of design that Bygate employed. It does not provide data that can easily speak to the effects of task planning on the acquisition of specific linguistic features. That is, it can only provide evidence of general linguistic change, as in types (2) and (4), but not of specific linguistic changes, as in types (1) and (3). To obtain evidence of the effects of task planning on specific linguistic features it is necessary to target specific features for study. This cannot be readily achieved by means of the kinds of unfocused tasks that have figured in task planning research to date. However, it may be achievable through the use of focused tasks. Whereas unfocused tasks allow learners to choose from a range of forms focused tasks aim to induce learners to use specific forms. In Skehan's (1998b) terms they are 'structure trapping' in that they make the employment of the specific forms, natural, useful or, ideally, essential (Loschky & Bley Vroman 1993). The advantage of such tasks is that they allow researchers to construct pre- and post-tests to mea-

sure whether learners knew the targeted forms prior to performing the task and what the effect of planning tasks is on learning. The only planning studies that have investigated specific linguistic forms to date are Ellis (1987), which targeted past tense forms, and Hulstijn and Hulstijn (1984), which targeted word order rules in Dutch. Somewhat disappointingly, more recent studies have been based on unfocused tasks.

Investigating learners' planning strategies

In a typical task planning study, learners are asked to carry out planning in accordance with instructions. Below, as an example, is the description of the 'guided planning – content focus' condition in Foster and Skehan (1999):

The students were introduced to the idea of a balloon debate. The teacher then presented ideas that each character might use to defend his or her right to stay in the balloon and students were encouraged to add ideas of their own.

Here is a description of the unpressured on-line planning condition in Yuan and Ellis (2003):

The on-line planners were required to tell the story by producing at least four sentences for each of the six pictures after seeing the pictures for only 0.5 seconds. They were given unlimited time to enable them to formulate and monitor their speech plans as they performed the task.

Such instructions raise a number of methodological issues. The most obvious one, given the evidence that pre-task and on-line planning have been hypothesized to have somewhat different effects on learners' performance of a task, is the need to ensure that learners receive instructions relating to both kinds of planning. In the case of studies investigating pre-task planning this has not usually occurred. That is, the learners are given instructions relating to how to conduct strategic planning/rehearsal but are left to perform the actual task in any way they choose. Thus, it is possible that the learners interpret the task performance conditions very differently, with some engaging in unpressured and others in real-time on-line planning. This may be one explanation why studies of pre-task planning have produced such mixed results for accuracy (see previous section).

There is also an obvious methodological need to establish whether learners actually carry out the planning instructions they were given. That is, do they conform to the prescribed planning conditions? Again, few studies have attempted to establish this. However, more recently, a number of researchers have attempted to describe the different strategies learners actually use during

the pre-task planning phase of a study. All three studies in the section dealing with pre-task planning in this book (Section 3) do this. The data used for such an investigation includes the actual notes that learners make while planning strategically (see Ellis & Yuan 2004) and post-task interviews with individual learners (Ortega 1999: Chapter 3 in this book). Such research is important not just to ensure that learners plan as intended but also because it can serve as a basis for drawing up guidelines for the design of effective planning instructions. Sanguran, in Chapter 4, makes a useful advance in this direction by formulating an explicit set of assumptions that guided her in the preparation of the planning instructions she used in her own study.

Somewhat different kinds of evidence are needed to demonstrate what kind of planning – pressured or unpressured – learners engage in on-line. While it may be possible to establish this through post-task interviews (although learners may have difficulty remembering their on-line decisions even if stimulated recall techniques are used), clearer evidence may be forthcoming by inspecting the fluency properties of the texts learners produce as a result of performing the task. Yuan and Ellis (2003) considered two such properties – the number of syllables produced per minute and the number of pruned syllables per minute (i.e. after dysfluencies had been discounted). They were able to show that learners in the unpressured on-line planning condition spoke significantly more slowly than learners in the pressured on-line planning condition. In this way, they were able to demonstrate that the unpressured on-line planners had performed as required.

Measuring learner production

Learner production can be measured either by means of external ratings or by means of discourse analytic measures. In general, language testers have preferred the former and SLA researchers the latter.

External ratings are based on scales that specify (1) the specific competency being measured and (2) levels of performance for each competency (often referred to as ‘bands’). In the case of ratings of task-based performance, the target competency can be specified either in behavioural terms that reflect the degree to which the learners have successfully completed the task (see, for example, Norris, Brown & Hudson 2000) or in linguistic terms. In the case of the latter, learners’ linguistic competency can be described either holistically (e.g. for the highest ‘band’ the descriptor might be ‘speaking proficiency equivalent to that of an educated native speaker’) or an analytic measure, where different dimensions of performance (for example, fluency, complexity and accuracy) are rated

separately. In Ellis (2003:298–302) I summarise the various options relating to external ratings.

In the case of discourse analytic measures, two types of measures are possible – measures of specific linguistic features and measures of general dimensions of oral and written discourse. There are a variety of well-established methods for deriving measures of specific linguistic features (e.g. error analysis, obligatory occasion analysis, frequency analysis and form-function analysis – see Ellis and Barkhuizen (2004) for a detailed account of these methods as they have been used in SLA). In the main, however, researchers have not used these, preferring instead general measures of learner production.

These general measures have been based on Skehan's model of L2 proficiency, which distinguishes two basic dimensions – meaning (fluency) and form with the latter further sub-divided into complexity and accuracy. Skehan (see Skehan & Foster 1997; Tavokoli and Skehan's study in Chapter 9 in this book) has been at pains to establish the independence of these dimensions by factor analysing scores obtained from a battery of measures. While the analyses do not always produce entirely similar results (e.g. in Skehan & Foster 1997 the analysis resulted in three distinct factors easily identifiable as fluency, complexity and accuracy while in Tavokoli and Skehan the analysis produced a somewhat different set of factors – temporal aspects of fluency, repair fluency and complexity/accuracy combined) they do broadly confirm Skehan's model. Thus, the general measures employed by Skehan and his co-researchers, have an established theoretical base.

There are nevertheless a wide range of measures of fluency, complexity and accuracy to choose from (see Figure 5 for a summary of the various measures employed in the studies reported in the subsequent chapters in this book). In one respect this is useful as, arguably, multiple measures of each dimension may yield a more valid assessment than single measures. The downside is that when researchers differ in their choice of measures it becomes difficult to compare results across studies. Ideally, work is needed to establish measures that provide the most valid assessment of each dimension (using, for example, a factor analytic approach such as that employed by Skehan), which can then be employed across studies. It is also worth noting that it may prove necessary to develop separate measures for spoken and written production, most obviously for fluency. Most of the measures used to date have been developed for oral production, as this has been the focus of the bulk of the planning studies. However, Ellis and Yuan (2004) developed measures of written production and Wolfe-Quintero et al. (1998) offer a comprehensive list of measures of all three dimensions for writing.

Type of measure	Specific measure	Description	Study
1. Fluency	Production rate	The number of syllables produced per minute of speech/writing	Ellis and Yuan; Kawauchi; Elder and Iwashita; Sanguran
	Breakdown fluency	The ratio between number of words reformulated and total words produced	Ellis and Yuan
		Number of repetitions	Kawauchi; Elder and Iwashita
		Total silence	Skehan and Foster;
		Number of pauses greater than 1 second	Tavakoli and Skehan
		Number of filled pauses	
		Length of run	
2. Complexity	Syntactic complexity	Ratio of clauses to some general unit (e.g. T-units, c-units or AS-units)	Ellis and Yuan; Kawauchi; Elder and Iwashita; Sanguran; Skehan and Foster; Tavakoli and Skehan
		Length of unit (e.g. T-unit)	Kawauchi
		Number of subordinate clauses	Kawauchi
	Complex grammatical structures	Use of comparatives and conditionals	Sanguran
	Syntactic variety	Total number of different grammatical verb forms used in the task	Ellis and Yuan
	Lexical variety	Mean segmental type/token ratio	Ellis and Yuan
3. Accuracy	Overall grammatical accuracy	Error-free clauses	Ellis and Yuan; Elder and Iwashita; Skehan and Foster; Tavakoli and Skehan
		Error-free clauses of different lengths	Skehan and Foster
		Number of errors per 100 words	Sanguran
	System-based grammatical accuracy	Correct verb forms	Ellis and Yuan
		Past-tense markers	Kawauchi

Figure 5. Discourse analytic measures used in the studies reported in this book

A final question concerns the length of the learner texts to which the measures are applied. In many cases, researchers do not use the full texts produced by learners but instead elect to use only part of the texts, typically the first five or ten minutes. The problem here, as Skehan and Foster's chapter in this book indicates, is that planning may have a markedly different effect on the first few minutes of production in comparison with later. Learners may have difficulty sustaining careful formulation and monitoring over a lengthier period of time. Skehan and Foster's study raises the awkward possibility that the findings of the research to date, which have typically been based on relatively short learner productions may not be generalizable to extended discourse.

Conclusion

Task planning has proven a rich vein for empirical study, as attested by the large number of studies that have investigated this implementational variable (larger than have investigated any other task variable) and by the current collection of studies. Why has task planning proven such a fruitful arena for SLA research? Is it just another fad in SLA, like the error evaluation studies in the 70s and 80s, that will soon lose its attraction? I think not. First, the study of task planning, as I have tried to show in this chapter has a strong theoretical basis drawing on a set of constructs (controlled processing, limited capacity memory, focus-on-form) and a number of well-established theories of L2 use and acquisition. Research, such as that reported in the subsequent chapters of this book, can both draw on this theory and help to test it. In a sense, then, the study of task planning lies at the very centre of current research in SLA. Second, the research is of obvious pedagogical relevance. Planning, whether of the pre-task or within-task kind, is a variable that teachers can easily manipulate in their day-to-day teaching. While teachers should not look to research as the only determinant of lesson design they can certainly benefit from the insights and 'provisional specifications' (Stenhouse 1975) that the task planning research offers them. Thus, for both theoretical and practical reasons I expect task planning to continue to attract attention in the years ahead.

This book constitutes an advance on the research to date. It addresses a variety of issues, some previously examined, others new:

- the role of task rehearsal in helping learners to elaborate content and to integrate the different strands of their L2 proficiency;
- the actual strategies learners employ during pre-task planning;

- the way in which learners orientate to the opportunity to engage in strategic planning;
- the extent to which learners' attention to form and meaning can be manipulated through pre-task planning;
- the effect of different types of planning (pre-task vs. on-line; detailed vs. undetailed);
- the interaction between strategic planning (a task implementation variable) and task design features (such as the introduction of a surprise element into a task);
- the effects of learners' L2 proficiency on their ability to make use of the opportunity for pre-task planning;
- the relative effects of unpressured on-line planning on oral and written production in an L2;
- the extent to which learners are able to sustain the effects of planning on performance over an extended period of time;
- the effect of context (e.g. a language test) on task performance subsequent to planning;

The range and variety of these issues testify to the richness of task planning as an area of SLA enquiry.

SECTION II

Task rehearsal

The chapter in this section examines the effects on task-performance of having learners repeat a task – of what was called ‘rehearsal’ in Chapter 1. Bygate and Samuda’s paper is important both methodologically, theoretically and pedagogically.

As noted in Chapter 1, the bulk of the research that has investigated the effects of planning on task performance has examined learner productions in terms of fluency, accuracy and complexity. There is an obvious need to extend analysis to the macro properties of learner discourse. Bygate and Samuda show that one way of doing this is by examining what they call ‘framing’. This is a cover term for a heterogeneous collection of linguistic resources used by speakers to convey ‘perspective’ (e.g. the speaker’s attitude to what is being communicated) and to ‘preview’ (e.g. by providing an advance organizer of what is to come). In effect, framing fleshes out the bare factual bones of a discourse. The analysis of learner narratives they present in terms of framing demonstrates that this constitutes a significant addition to the tools in current use. Bygate and Samuda’s analysis also points to the value of combining group-based statistical analysis with a qualitative, case study approach.

Their chapter is important theoretically because it provides a thoughtful account of how different kinds of planning (strategic planning, on-line planning and rehearsal) contribute to task performance. Bygate and Samuda argue that rehearsal offers the learners certain processing opportunities not available in the other types of planning, in particular the ability to integrate their linguistic resources. Repeating a task enables learners to reorganise and consolidate information into a richer, discursively more sophisticated performance.

Finally, Bygate and Samuda suggest that rehearsal is a useful pedagogic procedure not just because of the opportunities it affords learners to develop their L2 discourse skills but also because rehearsal arises in naturally occurring communicative activities (i.e. it has situational authenticity). The challenge facing teachers is to introduce task repetition in ways that students will find motivating.

CHAPTER 2

Integrative planning through the use of task-repetition

Martin Bygate and Virginia Samuda
Lancaster University

Introduction

This paper addresses an intriguing language teaching and learning puzzle: how to lead students to integrate prior knowledge into performance. Associated with this is the question of how best to help them to identify new knowledge needed for their development.

It is generally accepted that learning involves restructuring (McLaughlin 1990; Skehan 1998b). The term relates to a number of distinct aspects of the learning process, characterised in Ellis 1990 as noticing, comparing and integrating. Noticing of new elements in the input will often signal a change in the perceptual processes; noticing will bring with it an interpretation of the new input, and a change in the interpretations of existing knowledge. However, material that has been noticed (whether through explicit instruction, or through other cognitive/perceptual processes), although it is in principle *available* to the learners, may not in fact be drawn on. That is, a common learning and teaching problem is to get learners to integrate knowledge that is available to them into their active language use.

Indeed, one problem with communicative teaching is that this integration can fail to take place: as long as the learners are able to produce language which will achieve their communicative purposes, they may not do the additional work needed to extend their active repertoire. (Brumfit's 1984 separation of accuracy work and fluency work also seemed not to attend to this issue). Some aspects of this problem have been addressed in recent work on pedagogic tasks in L2 learning (see for instance Samuda 2001). In this paper we explore the possibility that doing a communication task a second time can help learners to

achieve integration of what they already know into what they do. The principle underlying this exploration is that repetition of a task enables two different experiences of the same task demands. The differences between the two experiences are seen as being due to different states of knowledge on the part of the speaker, and capable of enabling change. The first encounter with a task is likely to be the more creative encounter: the learner has to respond to a new demand. This is likely to mean that the learner has rather a lot of new work to do: for instance deciding how to do the task, what messages to produce, and how to produce them. In comparison, on repeating a task, the learner has valuable experience to draw on: after all, s/he has already internalised the information content, organised it into communication units, found relevant language to convey the meanings, and pronounced it. Hence on the second occasion the learner is likely to be under less pressure than on the first encounter, provided of course that the task is performed under the same conditions (with no additional time pressure, for example). Because of this, it is likely that at the first encounter the learner is more likely to rely on the most automated aspects of his/her language, than at the second. In contrast, at the second encounter, the learner is not only cognitively prepared, but furthermore, her/his vocabulary and grammar (especially vocabulary) are 'primed', so that there is a chance that on the second occasion the learner will generate more sophisticated output. This might involve such things as providing more backgrounding, and selecting a wider range of ways of formulating the message. In other words, an initial encounter with a task can be seen as creating a holistic representation of the task, along with the experience of handling it in real time. This representation and the accompanying experience can be stored, creating a kind of plan which can be drawn on on a second occasion, enabling the learner to integrate a broader range of their resources into their performance - that is, to perform more adequately their competence as it were (Clark 1974).

Types of planning

In the literature, two types of planning, *strategic planning*, and *on-line planning* are the ones widely identified (see Chapter 1). In this chapter, we are making a case for seeing task repetition as a form of planning. We argue that it is the experience of processing the task as a whole together with certain elements of both pre-task and on-line planning that is important. In this paper we call this *integrative planning*.

Strategic planning

Strategic planning typically involves focussed or unfocussed instructions to students to plan their performance on an upcoming task, for 2–10 minutes (for example Foster & Skehan 1996; Mehnert 1998). Results strongly support the theory that planning does affect performance: amount and type of strategic planning has effects on performance, notably on fluency and complexity, (Crookes 1989; Foster & Skehan 1996; Ortega 1999) but also on accuracy (Ellis 1987); and different effects have been shown to occur with different types of task. The underlying theory has not been elaborated in detail, but the belief is that strategic planning reduces the processing load of subsequent on-line performance: speakers may have mentally organised the content; and/or worked on the formulation of aspects of the communication. This preparation is held in memory and enables learners to produce more complex messages, both in content and in form, to produce them more fluently and to be more accurate.

There are however some limitations in concentrating on this type of planning. Firstly, it isn't at all clear how this kind of planning can affect learners over an extended period of learning (see Skehan & Foster, this volume). We see the impact on a specific performance, but the connection between the reported effects for a specific task and longer term learning is yet to be theorised and researched. Secondly, although sometimes strategic planning does naturally occur before certain speech events, and is very frequent before writing, strategic planning seems untypical of many oral activities. Hence, while it has clear potential as a pedagogic device, it is not a target condition for normal everyday speech production. Students will usually need to be able to perform adequately without strategic planning, and not depend on having this facility. The third and perhaps most important problem is that although it is not clear how far ahead speakers can plan, in many speech contexts, the amount of discourse that can be pre-planned is bound to be limited. A speaker may be able to plan the rough content – and some expressions – for the first two or three minutes of talk, but it is unlikely that they would be able to map out much further ahead in any detail, mainly because of working memory limitations. In other words, the construct of strategic planning is unclear in terms of its functioning.

Another issue concerns the focus of strategic planning. So far, research results have generally shown that strategic planning influences fluency and complexity more than accuracy. Although the reasons for this are not clear, a probable explanation is that in strategic planning speakers are more likely to focus on the substance than the expression of their talk, leading to reduced focus on accuracy. That is, prior to a task, macro-planning is on the whole more

manageable and more productive than detailed micro-planning of utterances. Useful language is harder to predict and harder to keep in mind than a content plan. Hence speakers are more likely to use strategic planning time to 'boot up' reserves of ideas, but less likely to check whether they have all the language needed to express them. Given this information 'charge', speakers would then tend to get into more informationally complex talk than they would without planning, accounting for the increases in complexity. This would be consistent with the findings in Yuan & Ellis 2003 that their strategic planning group produced greater lexical variety, but less accurate grammar (Yuan & Ellis 2003: 23): lexical variety may well be pushed by the pre-loading of information content.

Regarding the impact of strategic planning on fluency, there are also distinct pressures which are likely to lead to increased fluency. One is that the pre-task marshalling of ideas is likely to reduce the incidence of on-line hesitation in finding them; and the second is that some speakers at least will be motivated to speak faster in order to avoid loss of planned material from working memory. If this surmise is correct, content planning would tend to lead to greater fluency at a likely cost to accuracy. Following this reasoning, both the pressure towards complexity, and the pressure towards fluency would each tend to derive from the same source – attention to planned content. By the same token both pressures could often be expected to lead towards an increase in errors (it is perhaps worth recalling that typical advice to elementary second language writers is to keep the message simple to avoid errors). In addition, we anticipate that the increased focus on content and fluency may well be at the expense of the speaker's capacity to explore their grammatical range and monitor for accuracy.

All this is not to say that strategic planning makes it impossible for speakers to activate relatively unused (or less automated) language. But the scope for doing this, and especially of remembering it at the point of need in the appropriate utterance, is likely to be limited. Incidentally, this view suggests a potential separation between lexical and grammatical processing, which can be related to the two kinds of planning. That is, within a strategic planning condition, speakers will be more ready to attend to vocabulary than grammar, in line with VanPatten's (1996) view that the listener privileges the lexical rather than grammatical elements of speech (an insight also developed for both production and comprehension by George 1972).

Hence although it is clear that strategic planning can be a valuable pedagogical resource, the research results suggest that the procedure may have introduced a bias into the processing of speech, which could be detrimental to the focus on form. Strategic planning is likely to bias towards macro-planning

and away from grammatical work. This bias does not work in favour of the integration of available resources into students' performance. That is, if our interpretation is correct, then strategic planning would not simply increase on-task capacity, as intended in the research paradigm – it may also bias how that capacity will be used. So, in summary, although strategic planning may help access general 'declarative' knowledge prior to performance (which is clearly desirable in terms of language development), some of the knowledge that is accessed may turn out to be either irrelevant or else forgotten when the speaker is engaged in producing specific utterances. To the extent that strategic planning helps to activate the learners' knowledge structures prior to talk, memory limitations may constrain the extent to which these are actually engaged during the talk itself. Rather, attention may focus on information content, which, if accessed, may bias against attention to form.

On-line planning

The second type of planning which has been researched has been referred to as 'on-line planning' (Yuan & Ellis 2003). This is defined as the kind of planning which occurs during performance. It consists mainly of processes of message conceptualisation, lexico-grammatical searches, and monitoring, all at the level of particular utterances – that is, at the micro- rather than the macro-level. As has been shown elsewhere in this volume (see Chapter 6), this dimension is operationalised through allowing students more on-line performance time, on the assumption that without time pressure, they will engage in more covert planning activities than students performing under time pressure.

In contrast to strategic planning, this type of planning is likely to tax working memory less, since it occurs during the planning and production of specific utterances. Hence it may be more open than strategic planning to the range of different types of operation which speakers might need to engage in during speech production. As Levelt's 1989 model (see Chapter 1) suggests, these operations involve speakers not only in creating plans, but also in monitoring them prior to production. Hence if on-line planning time is used, speakers are likely to be better able to attend to the conceptualisation, the formulation and even the articulation of their messages. So that whereas in the context of strategic planning a speaker is unlikely to be able to produce or to recall many detailed plans (for instance at lexico-grammatical or articulatory levels), in contrast, on-line planning time may help speakers do precisely this. It may also give them space to self-correct after production.

If speakers do in fact do this when given the time, and are not biased by the task or by other aspects of the conditions to focus more on one aspect of the processing than another, it is reasonable to expect that accuracy and complexity might both benefit through this form of planning. However, it is possible that providing learners with plenty of time could have a detrimental effect on fluency. And indeed, the results of Yuan & Ellis's 2003 study are largely consistent with this: their on-line planning group 'produced language that tended to be more grammatically accurate but less lexically rich' but with lower levels of fluency (2003:23).

There are some puzzles surrounding on-line planning. One is that it isn't clear what aspects of planning speakers will focus on. From a pedagogical perspective it is not obvious whether speakers can be induced to attend to one aspect rather than another, and if so, how. And in research terms, it is not clear how the focus of students' attention can be tapped into through research – how can we find out what speakers are actually attending to? (But see Chapter 7). And as with strategic planning, there is no sense of how this type of planning can impact on learning in the medium- to long-term. These problems may be tractable. Most important for our purposes here is whether on-task planning – in enabling learners to attend directly to the production of upcoming utterances – can also enable them to access their broader declarative knowledge stores.

In considering this, it appears that the strengths and weaknesses of strategic and on-line planning may be reversed. As we have seen, whereas strategic planning is likely to help speakers to prepare broader conceptual plans, and access mainly receptive language stores (and research results are broadly consistent with this view), its weakness is that whatever language plans might be accessed pre-task can turn out to be irrelevant or lost in the light of actual utterances. The opposite seems likely to be the case for on-line planning: this type of planning seems more finely tuned to the needs of specific upcoming utterances, but in this case broader knowledge structures or language knowledge that is mainly reliant on controlled rather than already automatic processes may not be accessible. And this is simply because they have not been previously activated. In other words the *integration* of passive knowledge (Meara 2004) with proceduralised knowledge structures seems problematic with each type of planning, but for different reasons.

One solution that suggests itself to this problem may be to combine both strategic and on-line planning. This paper however explores an alternative approach – that of using task repetition as a form of what we might call 'integrative' planning. Is it possible that task repetition is one way of bringing

together macro-plans, passive knowledge, and language production, in a way which goes beyond the contributions of both strategic and on-line planning?

Task repetition as integrative planning

By task repetition we mean ‘repetitions of the same or slightly altered tasks – whether whole tasks, or parts of a task’. This definition enables us in principle to study the impact of task-internal repetition, although the study to be reported below in fact focuses on whole-task repetition.

To explain the relevance of task repetition to the issue of planning, it is useful to start from a consideration of the processes involved in task performance. To do this we draw on Levelt’s 1989 model of speech production, which we have been working with for some time. Although this model is framed for first language speech, the issues we are discussing in relation to second language speech are essentially the same – see for instance de Bot’s 1992 adaptation of the model for L2 speech processing. According to Levelt’s model, talk is produced via processes of *conceptualisation*, *formulation*, and *articulation*. If speech consisted of the separate productions of single concepts, made up of 3–5 words, then we might be forgiven for thinking that this is a sequential process. However, clearly most speech involves us in producing utterances containing more than one concept, and typically concepts will engage us in several words at a time. Hence, in normal speech production, the three processes of conceptualisation, formulation and articulation are probably best thought of as overlapping and therefore having to operate concurrently, rather like a cascade – the cascade in simultaneous action on three different aspects of the production process. The simultaneity requires a degree of automation at all levels, especially the articulation level. This account implies firstly that there are at least three different kinds of process involved in speech production and secondly that in second language production, any one – or all three – of these processes can be a potential source of difficulty (there are other implications too, but these are not our concern here). A speaker may have difficulty sorting out the conceptual content; or in finding words to express it; or else in articulating the words, each with different implications for planning.

If we consider the role of planning in relation to these processes, we can note some differences between them. First, articulation. Although it is not entirely impossible, articulation is unlikely to be the normal focus of explicit planning. Rather the knowledge needed for articulation is seen as stored in plans of pronunciation procedures, which are built up through experience, and heavily automated. There will be some continual monitoring, but since

pronunciations are mostly automated, this is not thought to be a major drain on attention capacity. The situation with formulation is slightly different. Formulation concerns the accessing in working memory of lexico-grammatical units which are appropriate to the intended message. Clearly, like articulation, this accessing process is also likely to be subject to a degree of automation, lexico-grammatical selection being normally accessed very rapidly in response to prior conceptualisation. However there are limits to automation here. For one thing it is a common experience that some accessing can require a longer search – some words or phrases or propositions are relatively rarely used, or entirely novel to the speaker, so require more time. In addition, some grammatical markers, which are also selected during the formulation process, can require more grammatical work than others, again implying that speakers may benefit from some freeing up of capacity, especially in the L2. Further, the formulation process will also involve some significant monitoring to ensure that the ‘draft’ formulation is appropriate. Hence if we compare the articulation and formulation processes, the former tends to depend very largely on pre-stored plans, which are less likely to be significantly affected by production pressures – either the plans are pre-stored or else they are not, and if they are not they are unlikely to be planned on the hoof. Additional planning time is less likely to make a difference. In contrast, the formulation process, being more open to judgment (Levelt 1989), is much more likely to be affected by some degree of on-line capacity, on which various types of planning could therefore have an impact. On the other hand, during conceptualisation, the speaker plans the content of her/his messages, both at macro level (the overall content of the discourse) and at micro-level (the conceptual content of each utterance). Someone who has a lot of work to do to sort out what to say, in what order, and how to connect it up is clearly likely to find that pre-planning time would help performance (as we have seen above).

How do these types of planning demand relate to task repetition? Performance of a task is likely to be affected by whether or not the speaker has previously performed it. Prior performance can be taken to mean that the speaker has already carried out a lot of relevant conceptualisation work, formulation work and articulation work. How might this affect subsequent performance? Given that articulation plans are already expected to be pre-stored, prior experience of a task is unlikely to affect the articulation process. In contrast, it can be expected to have a significant impact on both the conceptualisation and formulation work that has to be done during a re-run of the task. For the conceptualisation work, speakers can be expected to be able to use a re-run to easily bring back to working memory a trace of the whole task content. In addition,

details of the input which may have been lost from working memory on the first run are likely to be more easily noted on the re-run (familiarity with the input material at time 2 releasing capacity to notice more of the material than first time around). In a sense, this can be seen as enabling fine-tuning of the schematic memory store. In addition to this, conceptualisation is likely to be much quicker second time around, since much of the work has already been carried out.

For the formulation work, prior experience can be expected to have a number of influences on processing. First, links between conceptual content and lexico-grammatical forms have already been primed in the speaker's mind. This can have two effects – firstly to speed up the recovery of those links; and second to increase the capacity for lexico-grammatical searches to improve on previous formulations. Hence on the second occasion, formulation is likely to be speedier and more accurate. In addition to these influences, clearly the improvement of speed and accuracy of the conceptualisation processes outlined above is likely to make more capacity available at the formulation level. The speaker may have more space to look for words. Speakers might also make use of this spare capacity to monitor their formulations more thoroughly. In sum, the impact of task-repetition on formulation might be usefully described as one of 'integration' of potential resources into the actual performance of the task: that is, 'integration' in terms of the amount of content detail expressed, the speed of lexico-grammatical accessing, the appropriacy of lexico-grammatical selection, and grammatical accuracy. If we think of repetition as enabling a second 'draft', then task repetition involves targeting improvement not just of the draft (i.e. the language produced) but of the actual drafting process. That is, task repetition can have an impact on the processing, and not just on the product.

To summarise, repetition is theorised as having two distinct phases: a first enactment of a task, in which learners are likely to organise the cognitive content, scope out the likely useful lexico-grammar, and process it in real time, generating an experientially derived multi-level schema to support subsequent linguistic work; followed by a second enactment, during which the speaker can build on the previous one. In this sense, the initial enactment of a task is seen as a form of planning of processing and of content. We see this as having the potential to lead to integration of knowledge and performance, and as facilitating changes particularly in the conceptualisation and formulation phases of the production process. We would also anticipate that this would have an impact on the extent to which the discourses are elaborated.

Most previous studies have used general measures of fluency, accuracy and complexity (e.g. Skehan & Foster in Chapter 7). One (Bygate 2001) showed an impact of task repetition on fluency and complexity, attributable to increased capacity at time 2. Another study (Bygate 1996) also suggested an impact on the quality of the talk. Lynch & Maclean (2000, 2001) have shown similar effects on students' formulation. The present study explores extracts of the 2001 data set to focus on changes in the quality of the talk. In particular, we are interested in differences in the elaboration of the basic information content in the students' two enactments of the discourse, and whether any differences can be related to changes in underlying discourse production processes. We are also interested in whether this might be significant for language learning.

Method

The study we now report compares the oral discourse produced on two enactments of the same task, at an interval of 10 weeks, by non-native speakers of English (see Table 1). It aims to explore the impact of task repetition on framing (explained below). The study is in two parts. First it presents a statistical analysis of the use of framing in the speech of 14 non-native speakers in order to look at group trends. This is followed by a case study of three members of the group. The data are part of a larger database of recordings by 48 students in which speech produced on a repeated task was significantly more fluent and more complex than speech on comparable new tasks (Bygate 2001). In that study repetition of the same task was associated with differences in general aspects of speech production. This study aims to explore the quality of the language produced by focussing on the extent to which students' use 'framing' in their narratives.

Participants

Fourteen participants were randomly selected from a larger data set (see Bygate 2001). The participants were English medium students at a British univer-

Table 1. Design of the study

N=14	Task
Time 1	Cartoon narration
Time 2	Cartoon narration (+10 weeks)

sity. They had different levels of proficiency, and came from different language backgrounds. All were recorded narrating to a listener a short video-extract from a Tom & Jerry cartoon on two separate occasions. The two narrations were recorded 10 weeks apart. Different participants had different cartoons (four different cartoons were used) so that similarities in the results across participants cannot be straightforwardly attributed to the specific video-extract that was used.

Procedure

The implementation procedures were carefully designed and rehearsed by a trained interlocutor, so that conditions for each narration were as close as possible to being identical for all of the participants, and on each occasion. The recordings took place in a small but pleasant, well-lit teaching room.

Students were asked to watch a short – approximately 2½ minute – extract from a cartoon video, and after they had seen it, were then asked to recount what they had seen to an ‘interested’ listener. The material was not part of class work, students had no preparation or pre-task introduction, no planning time was provided and there was no time-limit on either occasion. There was no taught follow-up to the first narration, and students had no expectation that they would see the same video-extract on any future occasion. In fact, the question in the initial study was whether 10 weeks would be too long a period for students to retain any significant impact from the first encounter. However, as reported in Bygate 2001, an impact was in fact found on fluency and complexity. Hence this study builds on that earlier study and aims to find out whether in addition any qualitative differences can be found in the students’ talk.

Analysis

As we have said, the discourse is analysed in terms of the incidence of *framing* in the talk. The reason for focusing on this feature is that it represents an aspect of the complexity of the *discourse* produced by the participants, as opposed to the formal language complexity which has been studied to date, and which can be expected to vary in relation to whether or not the speakers have had previous experience of the task.

In this study, we use *framing* as a term to refer to any language additional to the narrative content. Such language includes anything that indicates:

- a. the perspectives of the speaker (such as evaluations); of the listener (such as the listener's likely interpretation or expectations); or of the characters (such as the characters' intentions or feelings, reasons, or their interpretations of what could happen, is about to happen, or has happened), and,
- b. previews to what is about to happen, abstracts of the whole story, summaries of the story, cohesive links highlighting the nature of the connections between parts of the story, explanatory background, foregrounding or highlighting.

We see these kinds of discourse features as framing because they are additional to the bare bones of the narrative. Given this, framing is clearly partly associated with backgrounding (Hatch 1983); however, as can be seen from the range of discourse features outlined above, we here conceptualise framing as a form of discursive contextualisation that can also provide an interpretative gloss on both backgrounded and foregrounded elements in the narrative, embedding as it were the basic narrative content into a texture of relationships between actors, actions, and the narrator.¹ Thus 'framing' is a very broad category, but one that is deliberately so since our interest here is in capturing a constellation of features that might work together to add coherence to the narrative.

The focus of our inquiry is whether or not 'framing' is more in evidence in one performance than the other. In operational terms, the analysis focuses on the presence of framing in one narration that is absent in the other. The null hypothesis is that there are no differences in framing in the two productions. The alternative hypothesis is that there will in fact be more framing at T2. This is of interest since it would suggest differences in speakers' conceptual processing, and in their capacities to handle the material.

Our research questions then are:

1. On re-enacting a task, are there differences in the amount of framing used by speakers on the two performances of the task?
2. Do any differences suggest that repetition made a significant impact?

Question 2 depends on demonstrating whether the changes are likely to be due to repetition, or whether they are due to other reasons, in particular whether any changes are attributable to overall language development resulting from 10 weeks of acquisition. That is gains in framing could be due to two main factors, overall lexico-grammatical development independent of the task, and greater mastery of information content arising from repetition of the task. Hence we also analyse changes in lexico-grammar. Evidence of significant gains in lexico-

grammar and information content would suggest that this could be the source for gains in framing.

It is worth noting however that the impact of lexico-grammatical development could be complex: lexico-grammatical development could also lead to gains in the expression of information content. But the reverse is equally possible, changes in the information content of the narrative could itself have an impact on lexico-grammatical performance at time 2. Hence significant changes in lexico-grammar at time 2 could have more than one source. A lack of significant change in lexico-grammatical production at time 2, however, would support the interpretation that changes in framing are mainly due to task repetition.

In what follows, we first of all report the statistical results for the 14 participants for the three measures (lexico-grammar, information content, and framing). We then consider how far the results suggest that changes in framing are attributable to repetition. We then turn to case studies of three of the students.

Transcripts were coded, and recoded by two raters individually. Inter-rater reliability was calculated at 90%, and problems were resolved individually on a case-by-case basis. As already explained, our scoring procedures aimed to identify and score all instances of framing, information elements, and lexico-grammatical features that were present in the students' narratives either at T1 or at T2. Thus a '0' score for a measure in one column means that there are no instances of framing, information elements or lexico-grammatical features in that performance that are different from those found on that student's other performance. Hence a score of T1 0; T2 0 would mean that all the items identified on the given measure were found in both performances, signifying *no change* from T1 to T2; a score of T1 0; T2 6 would mean that all the items found at time 1 were also found at T2, and that 6 additional items were found at T2, signifying a net *increase*; and a score of T1 6; T2 3 would mean that 6 distinct items occurred at T1 which did not occur at T2, and 3 distinct items occurred at T2 which were not produced at T1, (over and above those items which appeared at both T1 and T2), signifying a net *drop* in the numbers of items at T2. We wish to go beyond a statistical report however, and study some representative transcripts in some detail. This is on the grounds that the figures only make sense when related to particular student performances.

Group results

General statistical results

The group results are summarised in Table 2. The findings reveal overall increases from T1 to T2. As can be seen in the ‘totals’ row, aggregate and mean scores were higher at T2 for all three measures – for framing, information content and lexico-grammar. *t* tests show differences to be statistically significant for the first two measures. However results for the lexico-grammar measure are non-significant. We take this to imply that there is no significant change in lexico-grammatical proficiency over the 10 weeks.

On each measure, there were however overall mean gains of 3.4, 2.7 and 3.5 elements for Framing, Information Content and Lexico-grammar respectively. The non-significant result for lexico-grammatical changes probably reflects the considerable within-group variation. To get a better sense of the size of the lexico-grammatical changes from time 1 to time 2. We then calculated the proportion of the distinctive lexico-grammatical features at times 1 and 2, the size of the change from time 1 to time 2, firstly in terms of the total size of the corpus of 14 students, and secondly in terms of the total number of distinctive lexico-grammatical features at times 1 and 2 together. The results are shown in Table 3. As can be seen, the aggregate of words and grammatical features used exclusively either at Time 1 or Time 2, that is, the features which are being changed, amounts to 7% of the total number of words in the corpus. Of these, 3.3% (or slightly under half) occur at time 1, and 3.7% (slightly over half) at time 2. In other words, the overall lexico-grammatical changes from time 1 to time 2 amount to 0.4% of the total size of the corpus. This can be read in conjunction with the fact that the change is in any case statistically non-significant.

Framing results

As we have seen, there is a statistically significant difference in framing between times 1 and 2. Further, looking at the pairs of framing scores for individual participants (Table 2), increases are small for two students (4>5, and 2>3), but the remainder of the group show a substantial increase in distinctive framing devices (overall over 150%). So the answer to our first research question is, yes, there was a consistent difference in the amount of framing used at T1 and T2.

Table 2. Summary of group results

Student	Framing		Information		Lexico-grammar	
	T1	T2	T1	T2	T1	T2
AG	6	14	0	3	23	29
AS	4	12	4	7	33	41
BKR	4	5	10	4	30	19
CH	2	8	0	6	35	45
CN	2	3	1	8	15	33
DLP	7	10	6	1	36	24
EB	0	3	2	6	14	28
GA	7	2	1	5	28	20
IT	4	7	1	16	13	32
KW	3	6	3	10	18	25
LC	4	7	11	13	37	31
LL	8	11	5	10	50	35
UJ	5	12	1	6	24	37
ZD	4	8	2	4	23	30
TOTALS	60	108	47	99	379	429
Means	4.3	7.7	3.4	7.1	27.07	30.6
SD	2.2	3.7	3.5	4.0	10.56	9.14
Mean gains	3.4		2.7		3.5	
t-test results	p=.001		p= .008		n.s.	

Table 3. Proportion of lexico-grammatical changes

	T1 & T2 Distinctive lexico-grammar	Distinctive T1 lexico-grammar	Distinctive T2 lexico-grammar	Overall group gains
Total	808	379	429	50
% of total words	7%	3.3%	3.7%	0.4%

Impact of repetition on framing

With regard to our second research question, these results suggest that reworking the task did have a striking impact on the students' performance, in terms of the extent to which they frame the information they are using. In particular, the results seem indicative of underlying differences in the speakers' grasp of the material on the two occasions, and of the ways they may be processing it. First of all watching the video a second time is likely to be easier. After the second viewing, they are likely to have the information more clearly organised in their minds than on the first occasion. This in turn is likely to have an impact on their ability to think about and work with the content while talking. Clearer

Table 4. Students showing no lexico-grammatical gains

Student	Framing		Information		Lexico-grammar	
	T1	T2	T1	T2	T1	T2
BKR	4	5	10	4	30	19
DLP	7	10	6	1	36	24
GA	7	2	1	5	28	20
LC	4	7	11	13	37	31
LL	8	11	5	10	50	35

organisation of the content, and a greater capacity to think about it while talking are likely to affect the extent to which they are able to *frame* their talk. This then is our suggested explanation for the scores for framing at times 1 and 2.

The key question however is whether the changes in framing are attributable to repetition rather than to lexico-grammatical gains. As we have seen, the non-significant result for lexico-grammatical changes suggests that this is not a likely explanation for changes in framing. However given that there is evidence of variation within the group, in what follows we look more closely at the individual profiles.

1. *Students showing no lexico-grammatical gains*

Five students showed no lexico-grammatical gains, and these are presented in Table 4.

These five students actually used fewer distinctive lexico-grammatical features at time 2: although of course they re-used a lot of their language (implying no change), they reduced the size of their repertoire around it. For GA this seems to have coincided with a reduction in framing, but with an increase in information content. For DLP the reverse is the case – DLP reduced information content but increased the amount of framing. The remaining three students all increased the amounts of framing and information content, even though they reduced their lexico-grammatical options. The implication from this group then is that the incidence of framing was unconnected to changes in lexico-grammatical performance.

2. *Students showing lexico-grammatical gains*

The remaining nine students made overall lexico-grammatical gains. Their results are re-presented in Table 5.

As Table 5 shows, four of the students in this table show modest lexico-grammatical gains (AG: 6; AS: 8; CH: 10; KW: 7). The aggregate increase in framing and information content for each is more than the lexico-grammatical

Table 5. Students showing lexico-grammatical gains

Student	Framing		Information		Lexico-grammar	
AG	6	14	0	3	23	29
AS	4	12	4	7	33	41
CH	2	8	0	6	35	45
CN	2	3	1	8	15	33
EB	0	3	2	6	14	28
IT	4	7	1	16	13	32
KW	3	6	3	10	18	25
UJ	5	12	1	6	24	37
ZD	4	8	2	4	23	30

gains: AG: 11; AS: 11; CH: 12; KW: 10). Three of the students (IT, UJ, and ZD) show roughly the same number of lexico-grammatical gains as for framing and information content together, that is: IT 19:18; UJ 13: 12; ZD 7: 6. Given that some of the lexico-grammatical changes will have been due to improved lexical accessing, and some to random variation, we argue that it is extremely likely that at least some of the framing produced by these students is a product of familiarity with the task resulting from task repetition.

This leaves just two students who do show fairly substantial lexico-grammatical changes against relatively low gains in framing and information content. CN 18: 12; EB: 14 : 7. Profiles such as these suggest that in these cases the lexico-grammatical changes may account for some of the changes in information content and framing, bearing in mind however that here too some of the lexico-grammatical changes are likely to be due to improved accessing.

The picture overall, then is that for 12 of the students, gains in framing cannot be mainly attributed to changes in lexico-grammatical repertoire. Five of these, because they actually reduced the amount of distinctive lexico-grammatical expressions from T1 to T2; four, because although they increased their use of distinctive expressions, this increase was less than that for framing and information content; two because in their case their increase in lexico-grammatical expressions was at best on a level with aggregate gains for framing and information content. Only two cases raise the clear possibility of changes in framing being probably attributable to lexico-grammatical changes. Given these results, we conclude that the changes in the amount of framing used are largely attributable to factors other than to changes in lexico-grammatical performance. We believe that the main factor contributing to the changes in the amount of framing used is task repetition.

In what follows we now consider in detail the transcripts of three of the students discussed in this sub-section – that is, two of those who showed lower lexico-grammatical gains than gains in framing and information content (CH and AG); and one of those whose increases in lexico-grammar were roughly equivalent to gains in framing and information content (UJ).

Case studies

One of the advantages of considering grouped data as in the previous section is that it shows similarities and differences in the language use of different speakers, enabling a degree of generalisation, and making it possible to spot trends which might not be visible in a single sample. The disadvantage of course is that the numbers fail to show the quality of the language. They make it impossible to appreciate what happens to particular speakers, or what their talk looks like on the two occasions. Also, the statistical data fail to answer our second research question as fully as we would like. In particular a close examination of the individuals' language helps to show whether or not the increases in framing are associated with changes in language knowledge. Hence to show the kinds of language hiding behind the numbers, we present extracts from the transcripts of three of the speakers, in the form of small case studies. The full transcripts are provided in the Appendix, to enable readers to cross check our analysis against the full data for each speaker. Instances of framing occurring at one time but not at the other in a comparable sequence of the talk are coded in italic typeface.

Case study 1: CH

CH's statistical profile is shown in Table 6. To analyse CH's talk at T1 and T2, we compare two pairs of extracts from CH's recordings, one pair from the start of the narrations, and another pair from a little later on in the recordings.

We first consider the two T1 extracts (see Table 7). The left hand column presents the transcription, and the right hand column introduces a line-by-line

Table 6. CH's overall profile

Student	Framing	
	T1	T2
CH	2	8

Table 7. CH extracts – Time 1

CH: First narration	Analysis
<u>CH1.1 (START)</u>	
ah this was a film about ahm (,) a cat (,) and a dog and a mous:e	1. Framing via basic abstract
and the cat wanted to go fishing (2.0) hh (,) ah but (,) the mouse (,) oh he threw out the fishing line	2. Looks like framing abstract, but ‘threw’ suggests these are the first two actions
<u>CH1.2 (LATER)</u>	
an’ (2.0) hh ahm (,) he came up ou- (,) out of the water and the mouse (1.5) gave the cat a big (,) (tut) sot’v of baseba:ll (2.0) hm what’s it called? <a bat> baseball bat <yeah> hh ahm (1.0) to the cat (2.0) ah and (,)	3. Reports surfacing of the dog, and the transaction between the mouse and the cat. Speaker reformulation work.
the cat winded up the line and thought he had a catfish (,) but it was a dog hh	4. Reports actions of the cat, framed by the cat’s misunderstanding
and he started to beat (,) the dog with the bat just in order to kill him because he he still I think he still thought it was: a fish hh (,)	5. Reports actions of the cat, framed by account of the cat’s intention, and by cat’s misunderstanding.

analysis. Recall once again that any framing used here that does not appear at T2 would be marked in italics. As the reader will see, there are no cases of this in these extracts.

Most striking for our purposes is the speaker’s general focus on events rather than on reasons, intentions, interpretations, or cause-effect relations. The main emphasis in these extracts, then, is on observable events. Only four instances of framing occur: the initial abstract: ‘this was a film about a cat and a dog and a mouse’; and three utterances interpreting the cat’s thoughts or intentions: ‘thought he had a catfish’; ‘in order to kill him’; ‘because he he still I think he still thought it was a fish’. However, as the lack of coding shows, all of these four cases of framing also appear in the second narration. We note in passing that there is some evidence in the extracts that the speaker is having difficulty accessing appropriate words and structures, for example ‘baseball bat’. This, and the simple clause structure, together suggests formulation problems.

In comparison, the two corresponding T2 extracts, presented in Table 8, show more instances of framing than at T1.

At T2 (Table 8), the speaker provides additional framing throughout. Here we see a much more explicit opening abstract for the entire story, focussing on

Table 8. CH extracts – Time 2

CH: Second narration after 10 weeks	Analysis
<u>CH Extract 1 (START)</u>	
a:h i know this is hh <i>this is (,) tom having great fun (,) about (,) going (,) fishing (1.5) which jerry (,) effectively destroys hh</i>	1. Framing via abstract for the whole story
<i>as as: tom comes down to the beach and (,) with his rod and (,) and throws out the line</i>	2. Framing of upcoming action via explicit scene setting
<u>CH Extract 2 (LATER)</u>	
and then <i>in the end</i> when he has winded him on (,) <i>all the way into the beach</i> erm (2.0) <i>jerry provides tom with (,) a bat (,) to to (,) well immediately to to to (,) kill the (1.0) fish which is the dog (,) and erm (,)</i>	3. Framing via i) explicit marking of the end of an extended action sequence; ii) explanation for Jerry's action; iii) signalling T's misapprehension prior to the next action.
tom starts (,) beating (,) the dog (,) <i>not realising (,) that it is a dog until (,) it's too late! (1.5)</i>	4. Ensuing action, framed by explicit statement of the point when Tom's understanding changes.

the key elements of the story, the two key characters in terms of a typical 'Tom & Jerry' schema. This is followed by explicit scene-setting, showing Tom going down to the beach, the start of the whole episode. At the beginning of extract 2, the speaker provides a number of additions to the bare bones of the narrative: it is explicitly marked as concluding a durative sequence ('in the end' and 'when he has winded him on all the way...'); the speaker makes clear where this phase of the narrative is taking place ('into the beach'); the speaker expresses both Jerry's and Tom's motives ('jerry provides tom with a bat to kill the fish'); and expresses the fact that Tom is still under the impression that he is dealing with a fish when it is actually a canine ('to kill the fish which is the dog'). The final phase of action in this part of the transcript reports the beginning of the phase ('starts beating') and Tom's continuing misapprehension ('not realising that it is a dog until it's too late'). The Time 1 extracts run to 105 words, the Time 2 extracts to 89 words (including repetitions). Not only are the Time 2 extracts more economical, then, but they also provide more framing. So our argument here is then that at T2, CH provides a more *coherent* narrative, and one which is more schematised, in the sense that the discrete events which form the substance of the narrations are woven together into a rather more unified whole. At T1, in contrast, those events seem to remain somewhat separate.

We can now return to the question of whether the differences are due to changes in CH's language knowledge, or to increased processing capacity, and

whether this is attributable to familiarity with the task by considering the particular vocabulary items involved. If we consider those vocabulary items in these T2 extracts which do not occur in the T1 transcript, these are: ‘great fun’; ‘effectively’; ‘destroys’; ‘beach’; ‘provides’; ‘immediately’; ‘until’; ‘too late’. Did CH not have the resources to convey these meanings at T1?

We don’t know, of course, but can conjecture on the basis of the full transcript (see the Appendix). First, we note that at T1 the speaker uses phrases like ‘came up out of the water’, ‘winds back’, ‘winds up’, ‘round the foot of a big dog’, ‘weighted down’, ‘a fishing device’, ‘get hold of’. Given the reasonably sophisticated nature of these phrases throughout the T1 transcript, it is unlikely that CH didn’t know the words ‘great fun’, ‘immediately’, ‘until’, or ‘too late’. But if at T1 CH did *not* know some of the words used later at T2, alternative expressions would still have been available. For instance CH’s T1 word ‘smashed’ could have been used in place of the T2 ‘destroys’, and CH originally used ‘gave’ instead of ‘provides’. Looking at other T2 words, ‘effectively’ is not essential to this narrative, and ‘beach’ could have been replaced by some other word, such as ‘land’, ‘ground’, ‘floor’, or even omitted. Our conclusion then in the light of the student’s actual use of vocabulary is that the lack of framing at time 1 is unlikely to be due to a lack of relevant language resources. Rather there are grounds for believing that it is in fact a discourse issue, which we argue is more likely attributed to differences in processing capacity.

Case 2: UJ

UJ’s statistical profile (see Table 9) was very similar to that of CH above.

The video extract UJ was working with was different from CH’s. We consider two extracts from UJ’s T1 transcript, and compare them to corresponding extracts from UJ’s T2 version 10 weeks later. The extracts are from a point in the story, where the owner of the house can be heard coming downstairs, and Jerry gives Tom a kick so that he drops the plates that belong to the house owner. The T1 extracts are presented in Table 10.

UJ’s T1 performance consists essentially of seven simple actions, linked by ‘and’, with a final descriptive copula + adjective (‘happy’). This can be de-

Table 9. UJ’s overall profile

	Framing	
Student	T1	T2
UJ	5	12

Table 10. UJ extracts – Time 1

UJ First narration	Analysis
mm (2.5) after that hh er the mouse hh er hh (enough?) er kicks (,) the cat (,) er (,) an' an' the cat ah (1.0) oh let (,) let all the plates fall down and hh (,) mm: (2.0) and they all break (,) and (1.0)	1. A sequence of three actions, introduced by 'after that' and linked by 'and'
the: mm (,) mm the missus comes down and sees what happened an' (,) sentenced the cat seems to er strike her her it (,) with a with a oh hh (,) and throws it out an' (,) the mouse is happy to have the cat out of the house (2.5) ok?	2. A sequence of four actions linked by 'and' and one resulting state + focus, also linked by 'and'

scribed as a very bare narrative, with virtually no framing. In contrast in the second version (shown in Table 11), the speaker starts with a framing description of the scene, through an adverbial, and a progressive aspect of the verb 'hold'. This is followed by more framing through an adverbial phrase of purpose (marked by 'in order to'), which is then followed by two linked events. The second part of the extract similarly starts with a framing account of the context (an adverbial clause of time, plus a gloss about the house owner's interpretation of the situation). This is followed by the action, and concludes with the final framing summary.

The contrast between the two versions is encapsulated in the different clause function. In the first version the speaker uses seven narrative main verbs, and a single descriptive clause to conclude; in the second the speaker uses five narrative main verbs ('comes down', 'gives him a final kick', 'fall down', 'break', and 'throws the cat out of the house') but six framing finite clauses ('is still holding', 'comes', 'sees', 'makes ...responsible', 'has achieved', 'has the house free') and one framing non-finite clause ('in order to make Tom throw away the plates'). This is presented in Table 12. The table neatly illustrates the differences in balance between the two versions. As was the case with CH, at T2 UJ changes the balance to the point of devoting *more* talking time to framing than to the basic narrative.

Once again we consider the issue of whether the changes in framing could be due to changes in the speaker's language knowledge. Table 12 shows that there is little difference in the complexity or sophistication of verbs used. However words like 'responsible' and 'achieved' may give rise to doubts. Checking

Table 11. UJ extracts Time 2

UJ Second narration after 10 weeks	Analysis
<p><i>meanwhile</i> the missus comes down the stairs hh and (,) er <i>tom is still hh (,) holding all the plates</i> (1.5) mm hh in order to: (1.0) m- (,) make (2.0) <i>in order to make tom (1.5) throw away the plates</i> hh the mouse hh gives him a final kick hh a:nd (,) well all the (,) plates fe- (,) fall down and (,) break (2.0) ah s:-</p> <p>we ah <i>when the missus comes and sees</i> the (,) broken plates <i>she makes (,) tom (,) alone responsible for the (1.0) for the damage</i> and hh well throws (,) the cat out of the house hh ah so (1.0) <i>jerry has achieved (,) er er (,) its goal</i> hh and (,) <i>has the house free without the cat</i></p>	<p>1. Framing via: i) time coordination of new event; ii) description of T's position; iii) expression of purpose, which frames J's action, and resulting events.</p> <p>2. Framing via: i) time coordination of new event, reiterating an element of the previous episode; ii) metastatement about the landlady's interpretation, linked to resulting action, explicitly framed by iii) an evaluation of the outcome for Jerry, coupled with iv) an explicit statement of the final situation.</p>

Table 12. Cross-tabulation of UJ's use of narrative and framing verbs by time

	Time 1	Time 2
Narrative	kicks lets...fall down break comes down sees seems to strike throws...out	comes down gives him a kick fall down break throws
Framing	is happy to have the cat out...	is still holding comes sees makes responsible has achieved has the house free

back to the extent of UJ's transcript 1 (see Appendix), we note that at T1, UJ was already using a number of quite sophisticated expressions, such as: 'mocks'; 'sees the chance of revenge'; 'has to fetch'; 'otherwise'; 'fears that they might drop'; 'stack of plates'; 'throws it across the room'; 'the breaking of the plate'; 'comes to see what happens'; 'uses the tail as a towel'. This range of lexico-grammatical resources suggests that UJ's language knowledge at T1 was

Table 13. AG's overall profile

	Framing	
	T1	T2
Student	T1	T2
AG	6	14

certainly good enough to permit the kinds of framing we are considering. In other words, scrutiny of the transcripts suggests that the change in performance is not a matter of changes in available language, but rather *changes in attention*. To explore this issue over a longer stretch of transcript, we finally turn to the case of AG.

Case 3: AG

We first present the overall profile of AG's performance for framing at times 1 and 2.

AG's profile is similar to that of the previous two participants. However as we will shortly see from the extracts, like UJ, AG seems at a slightly higher level of proficiency than CH. The extract from AG's T1 performance is presented in Table 14.

Compared with the T1 narrations of CH and UJ, AG's T1 narration has more framing moves. This can be seen by comparing the analysis columns for the different learners at T1. However, if the repetition effect we are considering is consistent, it should operate irrespective of proficiency level. And indeed AG uses less framing at T1 than in the corresponding T2 narration. As can be seen from the lack of coded moves, most of the T1 framing moves also occur at T2. In contrast, when we examine the T2 transcript over the comparable stretch of discourse (see Table 15) we find significantly *more* framing devices. In addition, closer inspection also shows that the framing at T1 is consistently less explicit than at T2. For instance, in utterance 4, where the character's reasons are not mentioned; in utterance 6, where the speaker is not explicit about what the character has succeeded in doing; and in utterance 9, where the speaker fails to provide a complement clause to the adjective 'happy', leaving the focus of J's happiness unclear.

Equally interesting is the fact that AG's T2 version is much richer in framing, with multiple framing occurring in several of the utterances. While two of the eight T2 utterances are virtually identical to AG's formulation 10 weeks earlier, the remaining six utterances show far greater density of framing markers. For example, utterances 2, 3, 4, 5, 6, and 8 have 2, 2, 2, 5, 3 and 3 framing

Table 14. AG extract – Time 1

AG: First narration	Analysis
erm it was a comic strip about tom and jerry in the beginning (,)	1. Same as 10 weeks later
and tom (,) was erm: (2.0) holding (,) jerry by his tail and was teasing him (,) hh and erm (,) throw him (,) in the air	2. Framing via i) description of action, with ii) meta-description ('tease') followed by ambiguously unmarked verb 'throw'.
er jerry (,) could just manage to erm (,) erm (,) get hold (,) on a plate which was hh on a: kind of board up the wall kind of (1.0) gallery along the walls full of plates and (,) jerry could well (,) well (,) get hold there hh	3. Shift to unconnected action, speaker works to rephrase description of location.
and then jerry took revenge by throwing all (,) the plates (,) down hh	4. Framing abstract ('took revenge') plus repeated action, no account of motivation.
and erm: (,) tom was <i>desperately trying</i> to erm catch all the plates (,) jerry was throwing down	5. Framing via i) <i>account of Tom's state of mind</i> , and ii) intentions.
and erm hh (,) he finally (,) well (,) managed to do so	6. Framing of Tom's actions as successful without details.
leaning against the wall with all this high hill of plates (1.0)	7. Framing description of Tom's posture and location, including reference to the plates.
and there was only hh erm (,) one plate left (1.0) hh	8. Framing description, same as at time 2
<i>and erm (,) jerry (1.0) was very happy that he w- (,) ah (laughs) (3.0)</i>	9. Framing account of <i>Jerry's feelings</i> , without focus.
and finally jerry (,) throw the last plate down and it broke so tom was (laughs) (,) <i>very unhappy about that</i>	10. Framing of a two-event sequence via i) conjunct, plus ii) <i>account of Tom's state of mind</i> .
and the: landlady (,) of his house (1.0) she came (,) down the stairs <i>because she (,) had heard the noise</i>	11. Next action reported, <i>followed by a framing reason</i> .

markers respectively. That is, not only are there more framing markers overall, but some of the utterances are quite densely framed.

So what about the possibility that these changes are due to changes in the speaker's language resources? Although there are one or two words which AG

Table 15. AG extract – Time 2

AG: Second narration after 10 weeks	Analysis
erm it was a comic strip about tom and jerry (.)	1. Same framing opening as at T1
first (1.0) tom (,) is teasing jerry by (,) holding her by her tail and (2.0) throwing her into the air	2. Framing via i) <i>abstract 'is teasing' preceding account of actions, making its function clearer</i> ; ii) <i>'throwing' marked with framing function</i> .
nd then suddenly erm (,) when (laughs) tom (,) throws jerry into the air	3. Focusing via i) <i>time adverbials used to frame the critical event</i> ; ii) <i>the critical event itself marked as framing</i> . NB. 'throw' now marked as narrative.
she (,) lands on erm hh a board high up (,) the wall (1.0) and on the board there are (,) lots of plates (,) hh	4. Verb 'lands' i) <i>frames end of sequence</i> , followed by ii) <i>description of shelf framing the next action</i> .
so (,) erm jerry decides to turn the tables hh and starts to throw all the plates (,) down so that erm hh (,) er tom has to (1.0) catch them to avoid that they break hh	5. Framing via i) <i>explicit mention of J's decision</i> , ii) <i>aspectual 'start'</i> , iii) <i>explicit statement of effect ('so that')</i> , iv) <i>explicit modal ('has to')</i> , and v) <i>explicit statement of purpose ('to avoid ..etc')</i> .
and erm (3.5) he's doing so until he has a (1.0) big pile of plates hh and is (,) nearly unable (,) to to carry them (2.5) any more (1.5) so he's leaning erm against the wall (1.0)	6. Framing via i) <i>iteration of T's actions</i> ; ii) <i>linking 'until' to cumulative outcome ('holding a pile')</i> , iii) <i>consequences of the size of the pile – itself a critical condition framing the next action</i> .
and there is only one (,) plate left	7. Same information as at T1
so (,) hh jerry erm (,) throws the last plate down hh (2.0) and erm (1.5) hh (,) because she knows that er the (1.0) housekeeper (,) will come then and tom (,) is going to get some (,) problems hh	8. Framing of critical event ('throws') via: i) <i>introductory 'so'</i> ; ii) <i>reason explicitly stated in terms of J's inferred thought process</i> ; and iii) <i>the ensuing more important objective</i> .

may have added to his/her repertoire over the 10 weeks (for example, 'lands on', 'pile', 'glides'), there are few which the speaker was unlikely to have known at T1. Words such as 'suddenly', 'has to', 'nearly', 'unable', 'housekeeper', 'defend', 'drops', 'angry', 'takes', 'escape' are very unlikely to have been unknown at T1, given the fact that at T1 the same speaker uses words and phrases like: 'gallery', 'get hold of', 'desperately', 'punish', 'kicked in the back', 'took revenge'. Given this range of lexico-grammatical expressions that the speaker used at T1, al-

though it is possible that one or two new words were incorporated at T2, this comes nowhere near accounting for the sheer amount of framing which the speaker provides on the second occasion.

The results from the case studies, then, not only confirm that speakers used more framing at T2 than at T1, but also provide strong grounds for believing that the changes in the amount of framing can be reasonably attributed to their familiarity with the video clips, and to the likelihood that this altered their focus when producing the second narratives.

Discussion

In our view, then, the case study extracts provide a valuable perspective on what happens when speakers re-engage with the 'same' communication activity, and one that is consistent with the findings reported in Bygate 2001, and in Lynch & Maclean 2000, 2001. Our transcripts show that at T2 all three learners modify their productions in the same kinds of ways, irrespective of the differences between their T1 performances, or the fact that they were working with different cartoons. That is, while there are certainly individual differences between them in terms of the framing devices they selected, there are also striking similarities. In particular, the study shows that all three speakers increased the amount and specificity of the framing of their narratives. The impact was that their performances changed from initial disjointed reports of observed events to something much closer to an interwoven story, a very important qualitative change, suggesting a change from a recall exercise to the production of a schematised story. Further the case studies provide information which supports and helps explain the statistical results (illustrating incidentally, how quantitative and qualitative data complement each other).

We also argue that the transcripts provide support for the view that the differences between the speakers' T1 and T2 performances are not mainly due to changes in the speakers' linguistic resources. The consistency of the findings across the cases, and the statistical results from the analysis of the wider sample of students lends weight to this interpretation. In none of the cases are the lexico-grammatical changes sufficient to explain the increases in the amount of framing. Rather than differences in language knowledge, the changes seem to reflect a difference in what the speakers are doing, which can be traced back to the fact that they are already familiar both with the content of the task, and with handling it in real time. It is worth considering once again the specific ways in which the second encounter differs from the first.

Overall the speakers' work includes i) recalling; ii) creating a macro-organisation of the main chunks of the story; iii) micro-organisation of the bits of the story, in terms of foregrounding and backgrounding, and creating connections between the parts; iv) deciding how to formulate the meanings; v) expressing the formulations; vi) monitoring overall for accuracy. At T2 this work is clearly different. By then, the work involved in recalling and macro-organisation is likely to be much easier, given that the conceptual content is more familiar. This means that at T2 more information can be accessed. And since it was already organised once first time around, any organisation work this time is likely to be much lighter. This also means that work on the micro-organisation at utterance level is also likely to be easier, and formulation, expression and monitoring are likely to be more successful. The reduction of strain in these aspects of the task is likely to lead in turn to the speakers being able to attend more to the framing of their talk, and less to communicating the bare bones of what they can remember from their viewing of the video. Whereas first time around they are likely to be primarily focused on recovering enough content from memory, and putting it into comprehensible shape, second time around, speakers are likely to have more time and attention to give to the coherence of the story itself.

Other explanations for these results are of course available. However they are not persuasive. One is that the students practised their particular video extract. But this is highly unlikely: the students had numerous far more important things to attend to, including any English language programmes they were attending, the academic programmes they were following, and not least their social life at a UK university. What is more, since the data collection was quite unconnected to any English language programmes they were following, and the hypotheses being explored were not familiar to the teaching staff, any changes cannot be attributed to their specific language programmes. In addition, the students were involved in providing a number of other data samples over the 10-week period, and had no idea that in the 10th week they would encounter the same video extract that they had worked on in week 1. Furthermore, at week 10, data collection involved the collection of four distinct samples, of which this is just one. Given all this, it is implausible that the concentration of our fourteen students was targeted at retelling an account of a short video extract first seen 10 weeks earlier.

It is interesting to compare the changes reported in this paper with those that would be found under other planning conditions. That is, in place of task repetition, would similar effects be found with either strategic or on-line planning for a cartoon narrative? Although this is an empirical question, we

believe that there would have been some similarities and some differences, suggesting that task repetition is a type of planning, but one with its own contribution to make.

First the case of strategic planning. The option here would be to provide speakers with thinking time prior to the beginning of talk. We anticipate that this thinking time would have enabled speakers to sort through their memory of the video extract, and to forge some kind of a story from the elements. Doing this might well enable speakers to generate more framing than narrating without pre-planning time. Nonetheless we anticipate that with this type of planning, many speakers would begin their narratives without having resolved various aspects of the story they had watched. That is, their mental organisation of the story would not have been as thorough as in the task-repetition condition, since it is unlikely that strategic planning time would have enabled them to rehearse the whole story. As a result, in its telling the story would have been harder to frame: speakers would still be working to clarify the content. Furthermore, the strategic planning condition would have largely focussed on the content, and would not have taken the students through the various detailed lexical and syntactic searches which arise in the first telling of the repetition condition, and which provide preparatory priming for the second performance. Hence although some similar effects could be expected, we doubt that the speakers' control of the content or familiarity with the forms to be accessed would be as great as it is in the task repetition condition. The impact on their narratives would not be expected to be as great.

With regard to on-line planning, the comparison with this too is of course an empirical question. However we believe that on-line planning might favour speakers' lexico-grammatical searches, but without having already undertaken searches during a previous narration, they would not have rehearsed the particular lexical accessing. This would be expected to result in less successful searches in the on-line planning condition. Furthermore, without the 'settling' time allowed through strategic planning or task repetition, the content of the story would be less well organised, and the parts less well interrelated. Speakers would have to do this work during the on-line planning time. Hence we anticipate that the result would be less full and less effective framing. This comparison then may help to highlight the status of task repetition as one type of planning, with its own distinctive contribution to make.

Finally we consider two limitations in our study. The first limitation is that the category of 'framing' is rather broad. As defined here, it includes adverbial adjuncts, conjuncts and disjuncts, including adverbial clauses and phrases (notably of time, cause, reason, and purpose); aspectual markers of descrip-

tion; utterances with abstracting or summarising functions; modals; and verbs of cognition or intention. It could be argued that this grouping is linguistically too heterogeneous to hold any theoretical value. We would dispute this view. The justification for the choice of linguistic category is whether it is appropriate to the question being asked. Discourse complexity is bound to reflect the use of a number of different linguistic features. Hence unless there is a theoretically motivated reason for focussing in on the use of any one of the subcategories that are involved, it makes more sense to deal with them as a complex of categories. Indeed, in our view, this cluster of features all serve to embed the basic content into a more coherent narrative, and while each feature individually may be a fairly minor and even unimpressive indicator of change, the fact that taken together they collectively point in the same direction presents a more compelling argument. That is, it is far more likely that a 'framing' factor influences a number of low level features, rather than that it correlates with only one or two. There is doubtless scope however for exploring this further in future research, as well as for exploring how the broad constellations of features we have subsumed here under 'framing' relate to other categories of description.

The second limitation is that the present analysis does not draw on higher level discourse features. That is, changes in the amount of framing may not correlate with changes in the quality of the narration. Narrative quality is likely to be perceived at a more general level, rather than being directly reflected in the amount of framing (for instance there could be too much framing as well as too little, just as there could be too much information content). Consequently, although this study demonstrates significant changes in speakers' output, this doesn't fully account for changes in narrative quality. So although we argue that the evidence we have presented is strongly suggestive of qualitative changes in the speakers' narratives, fuller exploration of the issue has to be left to future studies.

Nonetheless, the analysis that we have carried out does suggest that the impact of repetition extends well beyond the domains of fluency, accuracy, and complexity, and into aspects of language use which involve qualitative issues such as how well speakers know the information that is being transposed into talk; how far speakers have already struggled to match language to content; and in what ways speakers bring their language knowledge into action to generate an effective piece of talk. The evidence and the accompanying argument in this section suggests that task repetition has a particular contribution to make in bringing learners to develop their grasp of what they are communicating, and their ability to exploit their language resources to communicate it effectively. If language teaching and language testing are concerned with the development

and assessment of discourse abilities, then it cannot neglect the kinds of evidence uncovered in this study. Teaching and testing discourse clearly requires attention to the conditions under which the discourse is produced.

Conclusion

Overall, then, the account emerging from the grouped and case study data suggests that speakers can use repeated encounters to ‘work’ with a task. It also suggests that in doing this they also work with the tools needed. Repeated encounters with a task make it possible for various processes to occur: information can be improved, reorganised, and consolidated; attention can be paid to different aspects of the narrative activity, such as to the way in which the speech event is formulated. That is, repeated encounters do not involve the learner in doing the ‘same’ thing, but rather in working differently on the same material. In classrooms this could be used – by learners and by teachers – to explore more fully what learners actually know, enable them to do things they weren’t accustomed to doing before, despite having the potential to do it, and could be used to identify key gaps in learners’ repertoires. In other words, repetition of some kind enables learners to work with a language problem in a reasonably stable site. In consequence, the types of learning we have in mind, and in particular the development of discourse skills, seem related to the effects of communicative repetition.

In terms of research methodology, the results reported here also suggest that categories of analysis can usefully be extended beyond the global measures of fluency, accuracy and complexity which have tended to be a standard in the literature, and incorporate measures which reflect the ways in which language is used to structure communicative events. In other words, this analysis suggests that understanding and assessing language performance on tasks really needs to take into account discourse complexity.

From a pedagogic perspective, we recognise that many teachers may be reluctant to ask learners to re-engage with a task they have already ‘done’ – the idea of task repetition does not at first glance sit comfortably with pedagogic principles that place a premium on variety and novelty. However, a task can be repeated in its essentials, without the speaker feeling that the whole speech event is pointless: we all commonly tell the same story to different people, ask different people the same questions, and sometimes return with good reason to say the same things to the same people. That is, the same material can be communicated more than once, and still engage the participants’ attention.

Here is perhaps an interesting pedagogical need – to find ways of maintaining the learners' interest while reusing the same material.

However in highlighting the benefits of task repetition as a form of integrative planning, we are not necessarily advocating a steady diet of repeated tasks in the classroom. Rather, we are more generally interested in the ways in which creative and communicative use of language occurs in the context of a range of types of communicative repetition, of which whole task repetition is one example. After all, repetition may involve all or part of the discourse structure, or the information content, or the interactional moves, or the lexico-grammatical formulation. The key issue is that it is not incompatible with creativity. And indeed, the evidence of gains in performance shown by this and associated studies argues that students can take advantage of an element of repetition to work creatively with the target language. Hence, while creativity and novelty are important in language use and in language learning, it seems likely that they also depend on an element of repetition.

What does this mean for the classroom? It is worth noting that there are already important elements of task repetition present in a lot of communicatively-oriented language teaching, although this may not be recognised as repetition *per se*. For instance, a number of well-known communicative activities are highly iterative in nature: 'mingling' activities, survey tasks and snowball discussions all involve learners in repeated engagements with the same content with different interlocutors. Repetition is also apparent in many pair and group activities, and, at more advanced levels, in activities such as simulations. Repetition, then, typically occurs in common learning activities. We suggest that teachers might – and indeed already often do – usefully exploit elements of repetition inherent in such tasks.

This study has illustrated some ways in which task repetition can combine and integrate strategic and on-line planning, albeit in slightly different ways. The findings suggest that in doing this learners do not necessarily engage in the kinds of strategic or on-line planning generally discussed in the literature; instead repetition seems to provide them with opportunities for strategic planning in the sense that they have completed one (or more) on-line performances (or rehearsals), and may of course also engage in on-line planning assuming they are under no time pressure.

It is of course true that we do not yet know what kinds of impact the various other kinds of repetition enumerated here actually have on performance, and this remains an intriguing avenue for future empirical study. However as this paper shows, we do now know something about the impact of whole task

repetition, which offers four reasons why repetition in general is potentially important:

- First, as this and related studies suggest, if repetition leads (or pushes) learners to work at relating the language they know – with its strengths as well as its weaknesses – to the content they are being challenged to handle, this is a valuable learning opportunity (see Samuda 2001 for an illustration of how a task can be used as both a starting and development point for the teaching of new language).
- Secondly, if it is the case, once again as suggested by this study, that repetition leads students to optimise their resources, and to perform at a more sophisticated level than they would otherwise achieve, then it is crucial that teachers and learners ensure that they are keeping in touch with the students' upper potential, and are not content to teach and learn at a level of operation inferior to the one they are actually capable of achieving.
- Thirdly, if as argued elsewhere (see Bygate 2004) repetition is a fundamental characteristic of ordinary discourse, then teachers should clearly consider building it into their classrooms.
- And finally, and perhaps crucially, it seems likely that repetition is an essential basis for students and teachers to be able to assess their language development, both on a lesson-by-lesson basis, and also across longer periods of instruction. That is, while the evidence of this paper is that repetition provides the students with in-built planning, it also provides a context for students and teachers to plan their subsequent language work.

Overall, then, we argue that the study provides evidence that enables us to link the general gains from task repetition reported in previous studies to more detailed changes in the level of specific transcripts. We also conclude that the evidence suggests that there are good reasons for considering task repetition as complementing both strategic and on-line planning, and that it has a particular contribution to make in encouraging and enabling learners to do two pedagogically important things: first, to integrate both receptive and active knowledge of language; and second, to elaborate both their grasp of the content that is to be verbalised, and the ways in which the verbalisation is to be formulated. Both of these aims are, we believe, central issues in language education and language development.

Note

1. We might take adverbs here as an example: conjuncts drive the basic narrative forward by signalling relationships between major events in the story, and also elaborate on and enhance the narrative by signalling how the speaker interprets those events; disjuncts enable speakers to indicate their attitudes to those events.

Appendix. Full extracts of three case study participants

Coding: speech in *italics* = additional framing

Case 1: CH

First narration	Second narration after 10 weeks
ah this was a film about ahm (,) a cat (,) and a dog and a mous:e	a:h i know this is hh <i>this is (,) tom having great fun (,) about (,) going (,) fishing (1.5) which jerry (,) effectively destroys hh</i>
and the cat wanted to go fishing (2.0) hh (,) ah but (,) the mouse (,) oh he threw out the fishing (tape noise) line (tape noise)	<i>as as: tom comes down to the beach and (,) with his rod and (,) and throws out the line</i>
ahm <i>but the mouse sot'v (,) played a trick on him an' (,) hh (,) and took the line (,) and (2.0) <mm mm> winded it round the foot of a big big dog lying on (1.0) is that a bridge? not really a bridge but anyway (,) hh (1.0)</i>	hh jerry (,) catches (,) oh well jumps in to the water and catches his line (,) and swims where were the (,) to the jetty where this (1.5) hh huge awful mm i dunno what kind of dog it is bull dog maybe (1.0) lies (,) sleeping and (,) binds the line (,) around (,) this dog's (1.0) foot (1.5)
so hh (,) the cat thought he had got a fish: and he's starting winding back the line (,) but there was a dog on the line (,) <i>hh (,) and various things happened! (laughs) to this dog</i>	<i>with the result that (,) that (1.5) tom thinks he has got a fish (1.5) when he (,) really has caught this this huge (,) dog (1.5) and he starts to wind back the line (laughs)</i>
he he got all the bridge (,) sot'v with him (,) when he tried not to get down in the water and he got (,) smashed between two stones	and the dog (1.5) is forced down in the water and he (,) he ends up (,) taking the whole jetty with him and getting it (1.0) on his head (,) and he gets (,) stuck between two stones (,) poor dog (laughs)

an' (2.0) hh ahm (,) he came up ou- (,) out of the water and the mouse (1.5) gave the cat a big (,) (tut) sot'v of baseba:ll (2.0) hm what's it called? <a bat> baseball bat <yeah> hh ahm (1.0) to the cat (2.0) ah and (,) the cat winded up the line and thought he had a catfish (,) but it was a dog hh

and he started to beat (,) the dog with the bat just in order to kill him because he he still I think he still thought it was: a fish hh (,)

and suddenly he realised it was a dog and he got! really scared (,)

hh now (,) the: (,) mouse (1.0) took the fishing (1.0) that's not either a fishing line but anyway fishi- fishing device (,) hh (,) ran away with it (,) an' up i- up into a tree (,)

hh an' of course (,) the dog was quite mad with the cat and started chasing him (4.0)

and the mouse (2.0) er (,) actually helped the cat because he he (,) weighted down the line so the cat could get hold of it and then winded him up so that the dog wouldn't get him (1.5) hh (3.5)

all right yeah well (4.0) I don't remember any more (laughs)

and then *in the end* when he has winded him on (,) *all the way into the beach* erm (2.0) jerry provides tom with (,) a bat (,) to to (,) well *immediately to to to (,) kill the (1.0) fish which is the dog (,) and erm (,)*

tom starts (,) beating (,) the dog (,) *not realising (,) that it is a dog until (,) it's too late! (1.5)*

and then *he he just tries to (1.0) pretend like (,) it's (raining) (2.5) but (,) obviously (,) the dog is not gonna buy that*

so he gets furious with with the cat and starts chasing him! (laughs) and

then in the end hh jerry comes to tom's help (1.0) taking the fishing rod's (,) climbing up the (,) tree and letting the line down just in time for tom (,) to get it (,) and then (,) wind him up hh so the dog actually cannot eat him

Case 2: UJ

First narration

Ok it's a cartoon about (,) i know it tom and jerry and (,) hh (,) ahm:

at the beginning the cat (,) wants to: (1.5) ah (1.0) (laughs) erm (3.0) e:r the cat like always (1.5) (laughs) (2.0) ah m- mocks the er the mouse and (1.5) and well er

the mouse sees the chance of revenge hh (,) and (,) er gets up the wall and hh er throws plates and dishes er (,) down to the cat

Second narration after 10 weeks

(clears throat) the cartoon starts er (,) when tom erm (2.0) er mocks er the mouse and plays with the mouse and (,) er (,) er

but the mouse can (,) er escape the cat (1.0) and (,) gets up to mm a board (,) which is (,) on the wall underneath the ceiling hh where are (1.5) where *plates and cups and other dishes are stored* (1.0) and (,) the mouse (,) starts throwing these hh plates down

and the cat hh has to: fetch them *because otherwise* hh er the missus or what (1.0) hh erm (2.0) would sentence him because the plates are broken

and he hh ends up with a high (,) erm (,) stack of plates there erm in his hands and (,) mm (,) *fears that they (,) might drop* hh and then the mouse takes the last plate and throws it across the room hh and e:rm that (,)

and hh the missus er (,) hears the breaking of the plate an:d (,) comes to see what happens (,)

then the mouse er (,) hh goes on (,) er mocking the cat an' hh swims in her er (***) erm milk hh (,) erm played in a milk dish in the milk and (1.0) hh uses the tail as a towel (,) hh and (,) *the cat doesn't l- seem to like it* (,)

mm (2.5) after that hh er the mouse hh er hh (enough?) er kicks (,) the cat (,) er (,) an' an' the cat ah (1.0) oh let (,) let all the plates fall down and hh (,) mm: (2.0) and they all break (,) and (1.0)

the: mm (,) mm the missus comes down and sees what happened an' (,) sentenced the cat seems to er strike her her it (,) with a with a oh hh (,) and throws it out an' (,) the mouse is happy to have the cat out of the house (2.5) ok?

and er: the cat wants to catch them hh bec- the e:r hh *in order not to break* (,) the: plates (1.5) and goes on to ca- e:r (,) tom (,) ca- e:r catching all the plates er

until the last one (1.5) hh jerry takes this the last one hh and throws it straight across the room and (,) well *tom can't catch it because* he has all the other plates in his hands hh (,) and *so* the plate breaks

hh er the mistress HH (,) the missus hh hears the: (,) plate breaking hh and (,) comes down *because* well she doesn't want her plates to be broken (2.0) er erm (1.0)

but the ca- (,) *the mouse er (1.0) sees its chance to:* (,) mock to annoy the cat hh now and well gets down jumps the cat on the nose hh and (2.0) slips down her e:r the cat's back in the: (,) milk bowl of the cat hh swims around in the milk bowl (,) and uses finally the cat's tail as a towel (2.0) hh (2.0)

meanwhile the missus comes down the stairs hh and (,) er *tom is still hh (,) holding all the plates* (1.5) mm hh in order to: (1.0) m- (,) make (2.0) *in order to* make tom (1.5) throw away the plates hh the mouse hh gives him a final kick hh a:nd (,) well all the (,) plates fe- (,) fall down and (,) break (2.0) ah s:-

we ah when the missus comes and sees the (,) broken plates *she makes (,) tom (,) alone responsible for the (1.0) for the damage* and hh well throws (,) the cat out of the house hh ah so (1.0) jerry has *achieved (,) er er (,) its goal* hh and (,) has the house free without the cat

Case 3: AG

First narration	Second narration after 10 weeks
<p>erm it was a comic strip about tom and jerry in the beginning (,)</p>	<p>erm it was a comic strip about tom and jerry (,)</p>
<p>and tom (,) was erm: (2.0) holding (,) jerry by his tail and was teasing him (,) hh and erm (,) throw him (,) in the air</p>	<p>first (1.0) <i>tom (,) is teasing jerry</i> by (,) holding her by her tail and (2.0) throwing her into the air</p>
<p>and then suddenly erm (,) when (laughs) tom (,) throws jerry into the air</p>	<p>and then suddenly erm (,) when (laughs) tom (,) throws jerry into the air</p>
<p>er jerry (,) could just manage to erm (,) erm (,) get hold (,) on a plate which was hh on a: kind of board up the wall kind of (1.0) gallery along the walls full of plates and (,) jerry could well (,) well (,) get hold there hh</p>	<p>she (,) lands on erm hh a board high up (,) the wall (1.0) and on the board there are (,) lots of plates (,) hh so (,) erm</p>
<p>and then jerry took revenge by throwing all (,) the plates (,) down hh</p>	<p><i>jerry decides to turn the tables hh and starts to throw all the plates (,) down so that erm hh (,) er tom has to (1.0) catch them to avoid that they break hh</i></p>
<p>and erm: (,) tom was <i>desperately trying to erm catch all the plates (,) jerry was throwing down</i></p>	<p>and erm (3.5) he's doing so</p>
<p>and erm hh (,) he finally (,) well (,) managed to do so</p>	<p><i>until he has a (1.0) big pile of plates hh and is (,) nearly unable (,) to to carry them (2.5) any more (1.5)</i></p>
<p>leaning against the wall with all this high hill of plates (1.0)</p>	<p>so he's leaning erm against the wall (1.0)</p>
<p>and there was only hh erm (,) one plate left (1.0) hh</p>	<p>and there is only one (,) plate left</p>
<p><i>and erm (,) jerry (1.0) was very happy that he w- (,) ah (laughs) (3.0)</i></p>	<p>so (,) hh jerry erm (,) throws the last plate down hh (2.0) and erm (1.5) hh (,) <i>because she knows that er the (1.0) housekeeper (,) will come then and tom (,) is going to get some (,) problems hh</i></p>
<p>and finally jerry (,) throw the last plate down and it broke so tom was (laughs) (,) <i>very unhappy about that</i></p>	<p>then (,) erm (1.5) jerry runs down the p- (,) pile of plates hh and erm (,) glides down (,) tom's (,) back hh (1.5) until (,) er she lands (,) in a erm bowl of (,) milk (1.0)</p>
<p>and the: landlady (,) of his house (1.0) she came (,) down the stairs <i>because she (,) had heard the noise</i></p>	
<p>hh jerry came (,) down from the board (2.5) and erm (,) started erm tearing (2.0) erm (,) tom (,) hh by his (1.5) tail <i>while he was still desperately trying not to break all the other plates hh (1.0)</i></p>	

and erm (1.5) whi- while doing this she was o- oh (,) she was (3.0) sitting in a: erm (1.0) pot of milk and was (1.0) having a bath (,) in there hh and erm drying (,) herself with (,) tom's tail

and when she erm heard that the landlady was now really coming she kicked him in (,) the back and so all the plates fell down and broke hh (,) and erm (7.0)

and so well the landlady was of course quite upset and she erm (6.0) erm (,) tore (,) erm (2.5) well tore erm (,) tom by his tail well (,) *in order to punish him probably hh* and er (1.5)

j- jerry was quite well triumphant and hid in her (,) mousehole (1.0) and finally she fetched a (,) kind of erm signpost (,) put it put it in front of her door and (,) on the signpost it said erm home sweet home

and (,) she erm (1.5) takes a bath (,) in the milk hh (1.0) erm hh (1.0) and d- (,) dries herself afterwards with tom's tail hh because erm (1.0) she knows that he can't defend himself because he's carrying all those plates hh (,)

and (2.5) then (,) they both hear that the h:ousekeeper is actually (,) coming (1.0) *so* (,) jerry (,) erm: *gives tom (,) a final kick (,) in the (,) back so that he hh drops* all the plates and they break (1.5) and (,)

of course the housekeeper (,) is coming and she's quite angry (,) *with erm (,) tom* and is (,) and so she (1.0) hh erm takes him by his tail and tears [?] him (2.5) behind her (,) hh while erm jerry is (,) (laughs) (2.5)

jerry manages to escape and is hiding in her (,) little (,) mousehole (1.5) mouse (1.0) hole hh and erm (,)

finally *we see that* (,) erm jerry is fetching a (1.5) kind of (,) sign board (,) which says (,) home sweet home (,) and she puts it in front of her (1.0) mouse hole

SECTION III

Strategic planning

Strategic planning serves as a cover term for a wide variety of pre-task planning activity. This variety is evident in the time allowed for planning, what learners are asked to when they plan, how individual learners orientate to the planning activity and the effect of the learning/instructional context on how learners plan. The three chapters in this section of the book extend previous research on strategic planning by examining what learners do when they plan strategically and what the effects of different approaches to strategic planning are on task performance.

These studies highlight the importance of two principal variables in explaining how strategic planning influences task performance. The first concerns the focus of learners' attention during planning, specifically whether it is on form, meaning or a combination of the two. Ortega challenges the dichotomy of 'form' and 'meaning', arguing that what is important is attention to 'form-in-meaning' and suggesting that this is precisely what strategic planning helps learners to achieve when they perform a task. Sanguran's study lends support to Ortega's claim. She found clear advantages for those learners who were told to attend to both form and meaning. The second factor concerns the proficiency of the learners. Ortega speculates that strategic planning may be of greatest benefit to advanced learners. However, Kawauchi's study indicates that this may not be the case. He found that his intermediate level learners benefited most where fluency and complexity were concerned and the low level learners where accuracy was involved. The advanced learners gained the least from the opportunity to plan strategically. Clearly, further research is needed to examine to what extent and in what ways the learners' proficiency interacts with the opportunity to plan strategically in the performance of a task. In this respect, it is likely that a third factor (the complexity of the task itself) will need to be considered. Interestingly, in this respect, Ortega reports that some of the learners she studied did not find strategic planning of much value because the task they were asked to perform was not challenging.

A common methodological feature of all three chapters is a concern for the process of planning as well as the product. Ortega argues explicitly for a

process-product approach to investigating strategic planning. She used post-task interviews to collect information about what the learners in her study did; Sanguran asked her learners to make notes while they planned and hand these in before they performed the task; Kawauchi had her learners complete a post-task questionnaire. Another common methodological feature of these studies is that they all made use of narrative tasks.

There are also interesting methodological differences in the three studies. Whereas Ortega deliberately incorporated a listener into the procedures for her task on the grounds that this would make it more authentically communicative, both Sanguran and Kawauchi opted for monologic task performances on the grounds that these would demonstrate what learners were capable of doing independently. One study (Ortega) allowed the learners to plan freely, while the other two attempted to manipulate what the learners did while they planned. Sanguran devised planning activities that required learners to focus on meaning, form or a combination of meaning and form. Kawauchi examined the effects of planning activities that were based on writing out the narrative, rehearsing the narrative orally and reading a model narrative. Two of the studies (Ortega and Sanguran) investigated learners of relatively homegenous language proficiency, while the third (Kawauchi) studied learners with mixed levels of proficiency. These methodological differences point to the complex nature of strategic planning and subsequent task-performance. The studies testify to the need to investigate this complexity by systematically exploring the variables that contribute to it.

It is clear from these studies, however, that strategic planning is beneficial to at least some learners and for some tasks. This is pedagogically significant as teachers can easily incorporate the opportunity for strategic planning into task-based lessons. Teachers will need to decide (1) with which tasks to employ strategic planning (perhaps only tasks that are challenging to the students) and (2) which learners will benefit most from strategic planning (perhaps the less proficient students). Here there is room for teacher experimentation. Consideration also needs to be given to how students can be encouraged to pay attention to ‘form-in-meaning’. In this respect, Sanguran’s principles for designing planning activities and the examples of activities with different planning foci in the appendix to her chapter may be of assistance.

CHAPTER 3

What do learners plan?

Learner-driven attention to form during pre-task planning

Lourdes Ortega

University of Hawai'i at Mānoa

Introduction

Over the last fifteen years, task planning has become a burgeoning area of research within task-based language learning. An accumulation of studies has converged to the conclusion that providing adult language learners with the opportunity to plan before doing a task allows them to produce discourse of higher quality in the second language. The benefits have been shown to hold true in a variety of languages taught in a range of higher education contexts, including (a) ESL in the US, the UK, and Australia (Crookes 1989; R. Ellis 1987; Foster & Skehan 1996; Wigglesworth 1997); (b) EFL in Japan (Wendel 1997) and China (Yuan & Ellis 2003); and (c) several foreign languages, namely Chinese in Australia (Ting 1996), German in the UK (Mehnert 1998), and Spanish in the US (Ortega 1999). Across these contexts and languages, adult learners were able to achieve higher levels of fluency and linguistic complexity during actual task performance as a result of engaging in pre-task planning. What is still disputed, and what will no doubt generate much more research in the future, is: Where do these benefits come from?

In this chapter, I will explore the nature of the benefits afforded by pre-task planning through an examination of what learners say they do when they plan. In order to do this, I will present the analysis of a corpus of post-task interviews obtained as part of two earlier planning studies (Ortega 1995, 1999). In the year 1994, at the time I designed the two studies, there was a strong interest in investigating the linguistic products of planning. This was the focus of the two seminal articles by R. Ellis (1987) and Crookes (1989) that by that time

had become standard citations in the task-based language learning literature, and of the research program being developed by Skehan and Foster, whose first paper on this topic (Foster & Skehan 1996) I had read in manuscript form. Documentation of the strategic processes engaged during pre-task planning, however, had not entered the stage of scholarly discussions. Yet, surely it was important methodologically to document the fidelity of treatments (that is, whether participants engaged in the planning behaviors researchers expected them to engage), as much as it was essential theoretically to articulate the range of strategic, metalinguistic, and metacognitive behaviors that planning was posited to foster. For both goals, retrospective data seemed to be the only means to begin to do so. Encouraged by Graham Crookes, my mentor in both projects, to pay attention to ‘the view from the learner,’ I decided to include post-task interviews in my study design to document what these learners really did when they planned.

A decade later, as of this writing, it is somewhat puzzling to realize that SLA research in task planning has expanded into important theoretical areas (e.g., individual vs. group-based vs. teacher-based planning, Foster & Skehan 1999; form-focused versus content-focused planning, Sangarun 2001; on-line planning and strategic planning, Yuan & Ellis 2003), but interest in the strategic processes enabled by planning remains still weak. I hope this article will stimulate interest in undertaking a process-product approach in future research programs.

The chapter is organized as follows. I first provide some background for the two studies and I explain the methodology for eliciting and analyzing the interviews. I then present the results, pointing at the importance of individual differences, language expertise, and attention to language during pre-task planning. More globally, I contend that the findings underscore the relevance of both cognitive and social/affective dimensions of task performance to arrive at a fuller understanding of planning. I conclude by arguing that pre-task planning facilitates a learner-driven focus on form and by calling for future research programs that adopt a process-product approach and encompass both task and learner in the study of planning.

The two planning studies: Background and linguistic outcomes

Participants

The participants in the two studies were adult learners of Spanish at the University of Hawai'i. They were in a foreign language context with little access to the L2 outside the classroom. There were almost twice more female than male students in the sample, and their ages ranged from 18 to 46. Twenty-seven were undergraduate students and seventeen were doing graduate or postgraduate studies. I was interested in investigating learners of different language proficiency in the 1995 and 1999 studies. In order to achieve this, I followed the sampling strategy which Thomas (1994) calls institutional status: the 14 speakers who participated in Ortega (1995) were recruited from fourth semester Spanish classes, whereas the 32 speaker participants in Ortega (1999) were recruited from fifth semester classes and beyond.

This strategy appears to have been successful. I infer this based on clear differences in the levels of complexity and accuracy observed in the planned narratives that the two groups produced, shown in Table 1. Most striking is the difference in mean length of utterance: On average, the speakers in the 1995 study were able to utter three words in one stream of speech under a single intonation contour, whereas the speakers in the 1999 study were able to produce utterances that were more than twice longer. The accuracy measures reflect a similar language expertise difference. On average, the learners recruited from fourth semester Spanish classes were accurate 72% of the time in their suppliance of the definite article in the planned narratives (the cases requiring indefinite and zero articles were too few to analyze), whereas as a group the learners in the second study were able to use the full Spanish article system at 87% accuracy levels.

In addition, in the interviews learners offered reflections about their proficiency that support a language expertise differential between the two studies. All speakers were volunteers and evaluated their motivation for learning and speaking Spanish as high, and their language-learning abilities as above average. Among the 1995 speakers, however, remarks about feeling self-conscious when speaking in Spanish were frequent:¹

... when it comes to communication, I feel lacking [laughs]... quite a bit, although I'm sure everybody [in my class] feels like that. We don't do a lot of speaking and everybody feels very awkward when they speak. I think that if you are only doing it two or three hours a week [attending Spanish class] it's really difficult to feel comfortable speaking. [Learner 95009]

Table 1. Comparison of complexity and accuracy indices for planned narrative in the two studies

	Low-intermediate speakers (n=14)		Advanced speakers (n=32)	
	mean	sd	mean	sd
Mean length of utterance	3.20	0.89	7.92	2.07
Target-like Use of definite articles	72.00%	17.80		
Target-like Use of all articles			87.00%	8.00

Note. Mean length of utterance involved counting all utterances and all (target and nontargetlike) words per narrative and calculating their ratio. The utterance was defined, following Sato (1988), as speech bounded by pauses and a change in intonation contour. Target-like Use was calculated following Pica (1983).

In stark contrast, most speakers in the 1999 study commented on having past experience interacting in Spanish outside the classroom, whether in the US or abroad. Given this triangulated evidence for a language expertise differential, I will refer to the participants in the 1995 and 1999 studies as low-intermediate level and advanced level speakers, respectively.

Elicitation of the oral narrative tasks with and without planning

In both studies, I visited Spanish classes and recruited volunteers who signed up as pairs. When each volunteer dyad came to my office for their scheduled session, I would randomly assign one student to the role of the speaker and the other student to the role of the listener. Although they did not know it, I was only interested in the speaker and had decided that the function of the peer listener was simply but importantly to make the narrative tasks more genuinely communicative. I decided this influenced by work done by Yule and Brown (see Brown 1995; Brown & Yule 1983). The narratives used in the studies (adapted from Hill 1960) were simple but contained a complete story with a twist in the end.

Table 2 shows the overall task design in the studies. As part of the elicitation procedures for each of the narratives, speakers were given a strip of eight pictures and listened to a recorded version of the story in their L1 (English) before their retelling. Immediately thereafter, the speaker would plan and tell (or only tell) the story in L2 Spanish for their peer. In both studies, speakers were simply told to use their time to prepare for the story in any way they wanted,

Table 2. Main elements of narrative-plus-interview cycles in the two planning studies

Task implementation phase	Task design feature	Comments
Pre-task	Input for narrative elicitation: – Picture strip – L1 recording Pre-task planning: – Time – Writing	To ensure comprehension of story input and comparability of resulting narratives. Repeated measures design: All speakers experienced availability and lack of pre-task planning.
Task	Monologic oral narrative for peer-listener	Authentic listener made task have a genuine communicative purpose.
Post-task	Listener: story retelling in writing Speaker: Retrospective interview	Listener occupied doing their writing task while retrospective interview, conducted in L1, takes place in adjacent room.

so the condition was one of unguided or general planning. They were also allowed to make notes while planning, but were informed they would not be able to keep them while telling the story to their partner. During the pre-task planning phase, I observed the event from a corner of the room and took field notes on the speakers' overt behavior while planning. Immediately after each narrative was completed, the listener engaged in a writing task about the story their partner had just told them, while I interviewed the speaker (in English) in an adjacent room. Each experimental session consisted of two (Ortega 1995) or three (Ortega 1999) narrative-plus-interview cycles, all following the basic procedures depicted in Table 2.²

Linguistic outcomes of pre-task planning in the two studies

As summarized in Table 3, in both studies planning had clear effects on the fluency and syntactic complexity of speakers' narratives. Specifically, it allowed them to produce fewer disfluency markers and to deliver their speech at a faster pace and to pack more words and more ideas in a stream of speech. By contrast, pre-task planning fostered greater lexical complexity in the narratives produced by low-intermediate but not by advanced level speakers and, conversely, some accuracy benefits were found in the planned narratives for the advanced but not for the low-intermediate level speakers.

Table 3. Summary of results for linguistic effects of planning in the two studies

	1995 study	1999 study
Fluency	More dysfluency markers in the unplanned narratives	Faster speech rate in the planned narratives
Syntactic complexity	More words and propositions per utterance in planned narratives	More words per utterance in planned narratives
Lexical complexity	Higher type-token ratio in planned narratives	No differences in type-token ratio
Accuracy	No differences in accuracy of definite article use or subject-verb agreement	More accurate use of noun-modifier agreement in planned narratives but no difference in accuracy of article use

Methodology: Elicitation and analysis of interviews

My research goal in the interviews was to elicit metacognitive responses that provided insight into cognitive processes associated with speakers' strategic planning of the tasks. In addition, I acknowledged that language learners are "thinking and feeling people acting with purpose that is generated by what they see as significant and meaningful for them *as learners* in particular social and cultural contexts" (Breen 2001:178, emphasis in the original), and I was hoping the interviews would help me understand the participants' perspectives on the opportunity to plan before speaking in an L2. Accordingly, the interviews were approached as immediate retrospective accounts that tapped participants' long term memory regarding their prior task performance (Gass & Mackey 2000), but also as open-ended exchanges in a social setting in which the researcher conversed with the participants and asked them to interpret their cognitive behaviors (O'Malley & Chamot 1990; Wenden 2001). As is customary in semi-structured interviews, I developed an interview protocol but modified it during the actual sessions. For example, I asked follow-up questions, often making use of my observations and the field notes I took during the pre-task planning phase, in order to probe deeper into the speakers' responses. I also made every effort not to pose directive questions and to stimulate their accurate recall by using students' planning notes and the task's visual stimuli.

The resulting interviews, which were all audiorecorded and transcribed, produced accounts between 20 and 30 minutes per speaker (approximately 10 minutes of interview per task). Of the total of 46 interviewees in the origi-

nal studies, the results presented in this chapter are based on only 44 due to loss of data for two speakers who participated in Ortega (1995). The data were analyzed in a two-pronged approach including content analysis of emergent themes and application of a priori categories. This analytical approach evolved over three recursive phases.

In the first phase, a content analysis of the interviews from the 1995 study was conducted by both my colleague Megan Thompson and me.³ In a first pass through the transcripts, we independently reviewed learners' comments and identified emergent categories. For example, we made annotations on the transcripts such as *planning helps*, *communicative stress*, *sensitivity to partner*, and so on. The next pass involved comparing our annotations, solving any disagreements we may have, and grouping our annotations into more generalized themes. In the second phase, I first carefully reviewed the interviews from the 1999 study, looking for the same categories and themes that we had generated from the 1995 interviews, and allowing for new categories to emerge. I then did a second pass through both sets of interviews to double-check annotations and themes of the content analysis.

The third and final analytical phase was motivated by the research goal to describe the cognitive operations involved in strategic planning in ways that could be related to SLA theories of pre-task planning. To this effect, all interviews were coded for a priori categories of learner strategies developed by other investigators. I chose the taxonomies proposed by O'Malley and Chamot (1990) and by Oxford (1990) to guide me in this process. I first compiled all strategies from these two classifications into a coding scheme and applied it to a subset of the transcripts. This proved largely appropriate, although a few strategy types emerged from the data that did not seem to have a counterpart in the classifications developed by either Oxford or O'Malley and Chamot. (In these cases, new categories were added, and they have been marked with an asterisk in Table 4.) After this process of refining and revising the strategy coding scheme, I coded all interviews. Subsequently, I trained a second coder and calculated intercoder reliability on 10% of the data. The obtained simple agreement of 79 percent falls in the low end of the range found by O'Malley and Chamot (1990: 117, 125, 131). In their studies, these researchers occasionally opted for consensus after discussion (p. 135). In the present study, however, all strategy results are based on my codings because I considered them to be more trustworthy, given my intimate familiarity with the coding scheme and the data.

Results

In what follows, I first present the results of the strategy report analysis, focusing on what cognitive and metacognitive operations speakers reportedly used. I then go on to discuss the findings of the content analysis of the interviews, which include learners' perceptions of planning, individual differences in learner orientation to communication or accuracy, and possible differences in the utilization of planning caused by varying language expertise. I then examine results that speak to the question of learners' allocation of conscious effort and attention to meaning and form during pre-task planning as well as during task performance.

Learners' strategy use during pre-task planning

Table 4 shows the strategy types documented in the interviews, together with the raw number and percentage of participants who reported the use of each.

Table 4. Strategies reported by the L2 Spanish speakers

Strategy types	Total sample (n=44)		Advanced level (n=32)		Low-intermediate level (n=12)	
	Raw	Percent	Raw	Percent	Raw	Percent
<i>Metacognitive strategies</i>						
1.1. Advance planning	19	43%	17	53%	2	17%
1.1.1. Previewing*	7	16%	6	19%	1	8%
1.1.2. Selective listening*	17	39%	16	50%	1	8%
1.2. Organizational planning	30	68%	20	62%	10	83%
2.1. Directed attention	9	20%	8	25%	1	8%
2.2. Selective attention	12	27%	7	22%	5	42%
2.3. Problem identification	15	34%	14	44%	1	8%
3.1. Production monitoring	33	75%	24	75%	9	75%
3.2. Monitoring impact on listener	19	43%	15	47%	4	33%
3.3. Auditory monitoring	10	23%	10	31%	0	0%
3.4. Visual monitoring	7	16%	6	19%	1	8%
3.5. Cross-language monitoring*	4	9%	4	12%	0	0%
3.6. Style monitoring	5	11%	4	12%	1	8%
3.7. Double-check monitoring	2	4%	2	6%	0	0%
4.1. Performance evaluation	18	41%	15	47%	3	25%
4.2. Repertoire evaluation	10	23%	6	19%	4	33%
4.3. Ability evaluation	8	18%	5	16%	3	25%
4.4. Strategy evaluation	16	36%	12	37%	4	33%
4.5. Prognostic evaluation*	7	16%	4	12%	3	25%

Table 4. (continued)

Strategy types	Total sample (n=44)		Advanced level (n=32)		Low-intermediate level (n=12)	
	Raw	Percent	Raw	Percent	Raw	Percent
<i>Cognitive strategies</i>						
1.1. Writing for retrieval*	18	41%	14	44%	4	33%
1.2. Elaboration	10	23%	6	19%	4	33%
1.2.1. Academic elaboration	7	16%	3	9%	4	33%
1.2.3. Visualizing	7	16%	5	16%	2	17%
2.1. Writing for later recall*	10	23%	10	31%	0	0%
2.2. Rehearsing	21	48%	17	53%	4	33%
2.2.1. Mental rehearsal	9	20%	7	22%	2	17%
2.2.2. Subarticulatory rehearsal	5	11%	4	12%	1	8%
2.2.3. Reading rehearsal	6	14%	5	16%	1	8%
2.2.4. Selective rehearsal	3	7%	3	9%	0	0%
3.1. Highlight & postpone	4	9%	4	12%	0	0%
3.2. Make-up & stick to it	2	4%	2	6%	0	0%
3.3. Improvise if needed	4	9%	3	9%	1	8%
4. Avoidance	6	14%	3	9%	3	25%
5. Lexical compensation	28	64%	21	66%	7	58%
5.1. Approximating	22	50%	15	49%	7	58%
5.2. Circumlocution & synonyms	19	43%	15	49%	4	33%
5.3. Lexical transfer	4	9%	3	9%	1	8%
6. Translating	25	57%	18	56%	7	58%
7. Across-language analysis	1	2%	1	3%	0	0%
8. Outlining/summarizing	37	84%	27	84%	10	83%
9. Text enhancement	3	7%	3	9%	0	0%
10. Using available support	9	20%	9	28%	0	0%
<i>Social/affective strategies</i>						
1. Lowering anxiety	11	15%	10	31%	1	8%
2. Encouraging oneself	10	23%	8	25%	2	17%
3. Empathizing with the listener	23	52%	17	53%	6	50%

Note. Categories are based on O'Malley and Chamot (1990) and Oxford (1990). An asterisk indicates a new strategy type that emerged from the data.

Overall, the most frequent strategies, which were reported by half or more of all 44 interviewees, were: *writing / outlining / summarizing* (84%), *production monitoring* (75%), *organizational planning* (68%), *lexical compensation strategies* of several kinds (64%), *translating* (57%), *empathizing with the listener* (52%), and *rehearsing* (48%). These strategies point at the preponderance of retrieval and rehearsal operations during pre-task planning, and they also foreshadow two themes that will be important in the content analysis of the

interviews: learners' attention to form and their social/affective concern with the communicative context.

Table 5 displays the mean number of strategies reported per speaker. On average, in their post-task interviews each learner reported approximately 12 different strategy types, of which 5 to 6 were metacognitive strategies, 5 to 6 were cognitive strategies, and one was a social/affective strategy. The group difference between the low-intermediate and the advanced learners was not statistically significant when compared by means of a *t*-test ($t(42) = 1.57, p = .12$).

Three generalizations stand out in these results. First, cognitive and metacognitive strategies were reported to similar extents. Second, although the strategy of empathizing with the listener was frequently reported, reports of other social/affective strategies were very limited in frequency and range. And third, the differences in strategy use related to language expertise were not quantitatively large.

The balanced documentation of cognitive and metacognitive strategies in the interviews is unexpected, since many studies have found that learners typically report more cognitive and fewer metacognitive strategies (O'Malley & Chamot 1990). This suggests that the use of the metacognitive strategies by the L2 Spanish speakers was enhanced by the opportunity to plan prior to completing one of the narratives.

Table 5. Descriptive statistics for strategy types reported by speaker

	All strategies	Metacognitive	Cognitive	Social/affective
All participants (n=44)				
Mean	12.11	5.20	5.91	1.00
Median	12.00	5.00	6.00	1.00
SD	4.67	2.61	2.84	1.01
Min./Max.	4/22	1/12	1/13	0/3
Advanced level group (n=32)				
Mean	12.78	5.53	6.16	1.09
Median	12.50	5.00	6.00	1.00
SD	4.48	2.51	3.06	1.03
Min./Max.	5/22	2/12	1/13	0/3
Low-intermediate level group (n=12)				
Mean	10.33	4.33	5.25	0.75
Median	8.00	3.50	5.50	0.50
SD	4.92	2.77	2.09	0.97
Min./Max.	4/19	1/9	1/8	0/3

The limited range of social/affective strategies documented in the interviews, on the other hand, comes as no surprise and is likely an artifact of the methodology employed.⁴ Namely, reports on the use of affective strategies are elicited easily through questionnaire items but only infrequently in open-ended self-reports (see O'Malley & Chamot 1990: 222; Anderson & Vandergrift 1996).

Another consistent finding with the learner strategies literature is that differences related to language expertise are qualitative rather than quantitative in nature (Chamot 2001), and the present study is no exception. What is needed to complement the partial picture afforded by the quantitative analysis of strategy use, then, is a qualitative look at the themes that emerged from the interview data.

Learners' perceptions about planning

Learners' own perceptions of pre-task planning is an important piece in helping us understand how and why planning worked, and in what ways it may not have worked for everyone, at least not to the same degree. Speakers' appraisal of the putative benefits of planning fell into four patterns. Many learners (twenty-six or 59%) thought that planning had definitely helped them tell a better story and said they felt less stressed while telling it. A smaller portion (ten or 23%) stated that the availability or lack of planning time did not make a difference in terms of how they felt about task difficulty and success of performance. Finally, four speakers (9%) were ambivalent in their appraisal of planning, and another four (9%) indicated that they definitely preferred it when they did not have time to prepare for the task in advance. In Tables 6 and 7, I have summarized the perceived benefits and limitations typically brought up by the speakers.

As can be seen in Table 6, having extra time and being able to write notes were discussed as two distinct benefits afforded by pre-task planning. Overall benefits from having extra time were associated to being able to "collect one's thoughts" or "digest everything first" and to identify language problems ahead of time. The single most frequently mentioned benefit of having extra time, however, was the opportunity to engage in lexical searches. Apparently, these speakers felt time had enabled them to retrieve vocabulary that would have been otherwise inaccessible. By contrast, for many speakers it would seem that it was writing rather than extra time that allowed cognitive processes which (a) supported rehearsal operations, (b) encouraged them to complexify their story, (c) helped expand the range of lexical choices learners made by pushing them beyond "first year vocabulary," and (d) were the thrust for a focus on grammar (cf. illustrations in Table 6).

Table 6. Benefits of pre-task planning identified by learners

Beneficial feature of planning	Function	Example
Having extra time	1. To organize thoughts	I got to kind of digest everything first [99008]
	2. To formulate thoughts	I was just looking at the pictures trying to see what the important elements are and to think how to phrase those things [99005]
	3. To solve lexical problems	I figured out some of the words that I may not have been able to come up with right away and I figured out ahead of time [99004] I was able to rack my brain for some other way to say it [95013] I first went picture by picture to see where I was gonna have problems, then I tried to find a way around it. [99032]
	4. To practice/rehearse	I could play the Spanish in my mind [99028]
Writing notes	1. To formulate thoughts	I think for each frame I kind of jotted down- on my notes there, I was trying to think of the words that I would use, just various phrases that I ought to remember. "She puts her purse down" just remembering that, it's a key part of the story, I felt it was important. [+99010]
	2. To help lexical retrieval	Writing it helped me to remember how to say things [99002]
	3. To practice/rehearse	then I read it through, and mentally practiced [99024] In the end I was trying to see if I can remember everything without looking at the notes [99015]
	4. To improve overall content	I pulled out some extra details [99021] That [writing and rehearsing] gave me the chance to reach, you know, out to the side and point out to more things as I went along [during delivery of the story]. [99024] after I finished I went and added a few little details to each [picture], since I had the time [...] I was just trying to add little details, just in case I might need them [99005]

Table 6. (continued)

Beneficial feature of planning	Function	Example
Writing notes	5. To improve lexical choice	I wrote a couple of more advanced words [99021] I noticed that if I can write it out then I come up with a whole lot of words but then if I just have to tell and just say it right, then and there, then I limit myself to small vocabulary, like the vocabulary I learned from my first year of Spanish, that's my strongest, so I noticed that. [99006] [in the planned story] I guess I was trying to use more words, not just saying "and then" "and then", you know, that's why I said "all of a sudden this, that" [99008]
	6. To help grammatical retrieval	I was able to figure out which conjugations I could use [99002] I tried to write out the verbs and kind of the actions that were going on [95004] I could see where I was supposed to put articles and that stuff [99006] when I was writing it's like I remembered the subjuntivo, so I used it [99015]
	7. To help monitor grammar	I made little revisions, like <i>cumpleaños fiesta</i> {birthday party}, then I turned that around {i.e., <i>fiesta de cumpleaños</i> , Spanish word order}. [95011] I corrected those verbs, even though I'm not gonna have the notes to do the exercise, but it still helped me to know, that's how I remember things, by writing them, so. Cause writing, even though I never got to look at it again, it helped me to remember how to say things, and I always have a hard time with the conjugations. But I was able to figure out which ones I could use, so that was helpful. [99002]

In sum, the benefits summarized in Table 6 point at the centrality of retrieval and rehearsal operations. Retrieval operations, in particular, benefited organization of thought, access to a wider range of lexis and grammar, and elaboration of content and vocabulary. The link between retrieval and complexity is corroborated by the linguistic outcomes in both studies (cf. Table 3) and supports the conclusion, put forth by Crookes (1989), that one of the main

benefits of pre-task planning is that it allows learners to access the upper limits of their interlanguage grammar without time pressure, thus making a wider linguistic repertoire available for subsequent on-line use.

The benefits of planning may not have been reaped by everyone, at least not to the same degree. The eighteen (41%) speakers who did not necessarily perceive the availability of planning opportunity as advantageous gave varied reasons, summarized and illustrated in Table 7.

Table 7. Limitations of pre-task planning identified by learners

Locus of limitation	Limitation	Example
Performance conditions	1. Low task complexity	But your thoughts are already organized cause you've got to go in the order that pictures show, anyway [99014]
	2. Poverty of planning conditions	Ordinarily I would've had a dictionary, or a friend to ask, or something [99023]
Language expertise	1. Ceiling to retrieval benefits	Maybe I remembered a couple of words more, but then afterwards nothing more came [99026] What I couldn't remember, I didn't remember, still [99014]
	2. Lack of transfer to on-line performance	I forgot what I had practiced [99028] But when I said it on the tape I didn't say any of that. I stuck to just, I segmented them even more than when I was discussing it on my own. [95011]
Learner preferences	1. Added pressure with planning	A little bit more tense to do it, because getting to write the notes it meant "oh I have to remember more things". [99021] When you prepare you get worried. You get nervous that you are going to mess up and then you probably do. That's what I think. [95004]
	2. Extemporaneous performance feels good	When I have time to think about it, I get even more careful, even though it may be more proper, the language, I still feel that it's false, the talking [...] I just talked like I normally talk [in the unplanned story], it felt so much more natural. [99025] I would rather come up with it. I would rather decide what to say when I'm saying it, than to plan everything out ahead of time. [950013]

Two limitations related to the *performance conditions* were the inherent simplicity of the stories, which rendered the need for pre-task planning superfluous for some speakers, and the perception by two speakers that time or writing alone could not make a real difference without having other resources to plan, such as a dictionary or an expert user of the L2. Other identified limitations of planning seemed to have their locus in the learners' *language expertise* rather than in the external performance conditions. Specifically, several speakers in both studies stated that there was a ceiling to what extra time or writing could buy them in terms of retrieval, since "what you don't know, you can't remember." A lack of transfer of the plan to on-line performance was also reported by several other learners, who suggested they had focused on complexity or grammar during planning time, yet found out that they could not make use of what they had planned to say once they were retelling the story. Finally, other perceived limitations to the benefits of planning seemed to be related to *individual learner preferences*. For example, a few speakers mentioned that planning put a certain additional pressure on them by making them more accountable for their on-line performance. Several learners felt a certain pride in being able to perform well "off the top of your head" without having to plan first, and they articulated a personal preference for extemporaneous performance not only in the context of the study, but in real-world communication in general.

The themes in Table 7 hint at the moderating influence on planning of individual differences and language expertise. Inasmuch as these two factors influenced learners' perception and utilization of planning, they are potential areas of interest in future planning investigations. Therefore, in the next two sections, I explore these two areas in more detail.

Individual differences in learner orientation:

Communication versus accuracy

In a number of clear cases, the interviews revealed there was a natural divergence in task approach between some speakers with an overall inclination towards communication and some speakers who were more predisposed towards accuracy. Learners who I characterize as displaying a strong communicative orientation spoke of successful communication in real-world situations as self-regulated through the features of context and interaction:

When I speak in Spanish, especially when I'm kind of struggling somewhere, I'll just use any old word because I feel eventually in the course of the whole conversation people pick up whatever you're talking about, even if you use

the wrong word [. . .] if I'm talking to someone and *pelota* {ball} happens not to be the right word for football, they're gonna pick up in the context what I'm talking about and then they'll be correcting me midway down the conversation, they'd say 'Oh you mean this other word', and then I'd keep going, and you know still it's not lost, of course it's understood, because of that feedback your story is not lost, you don't have to go back to the beginning of your story. [99010]

Communication oriented learners seemed to accept error and error correction as inherent to their being non-native speakers of the language, and as part of a gradual process of second language learning. Moreover, these speakers may not have viewed accuracy as a realistic goal for L2 learning:

Unless you're a native speaker, you're not gonna be able to speak perfect, and you always miss some word here and you always have some drawback [...] I mean, we're still English speakers, you know? [...] Cause, I mean, you want to say it correctly, but I don't think you need to say it always correct, I think you can always get away with it, or get along the idea in a basic way, yeah, you want to say it correctly, but it'll come through repetition and going over more and more, or being among people that speak correctly all the time. [99013]

Another speaker explains how she does not worry about making mistakes or about being corrected, which interestingly she interprets as a sign of her maturing and becoming a better language learner:

I don't think that grammar is nearly as important as context, I used to think it was very much more important and I used to hate speaking in a foreign language and making a fool of myself, but I think I've gotten to the point now where I think it's more important that you communicate rather than you're perfect in your grammar and your pronunciation and everything else [Interviewer: How did you change your mind about it?] I think I just grew up a little and got less self-conscious [...] now I've been corrected enough times so it doesn't bother me [laughs]. [99002]

As mentioned in the previous section, speakers with a strong orientation towards communication expressed a definite preference for extemporaneous over planned performance (see Table 7), and thus they were skeptical of the advantages of planning.

The exact same range of issues appears to be conceptualized in a very different light by a few speakers who I characterize as oriented towards accuracy. They complained that it was frustrating not to be able to say what they wanted to say, and they expressed feelings of intimidation when facing the non-classroom demands of L2 communication situations. Two speakers in particular, [99005] and [99015], both advanced level learners, talked of their natural

predisposition towards accuracy in very candid ways. They were anxious about making mistakes, and they seemed to view L2 learning as a prolonged effort to reach “a hundred percent correctness”:

I always go back and think about the errors that I make, almost to the point where I don't want to say anything if it's going to be wrong, although I think I've gotten passed that problem, you see, otherwise you can't say anything, but yeah, I do want it to be perfect, I'm always trying to get it a hundred percent perfect, I'm never happy with just sort of less. [99005]

Even more extremely, learner [99015] remarked that speaking correctly could sometimes be more important for him than the quality of the content of the message:

[Interviewer: What would you say your focus was when you prepared the story?] What was my focus. It might have been that all sentences that I said were correct in Spanish, I think I paid more attention to that than to actually telling the story [I: How come?] I don't know, especially when I speak in another language, I'm always conscious of making sentences correct, even if I don't say anything worth listening to. [99015]

These accuracy oriented learners seemed to have a lesser concern for the impact of their performance on the listener than most other speakers (see discussion on learner sensitivity in a later section), and commented that allowing for planning is a question of academic fairness when stakes are high in the classroom (e.g., before an oral test).

The evidence for learner orientations I have presented in this section must be taken cautiously, because the original studies were not designed to address this issue. Thus, future studies may need to feature systematic questions so as to more precisely relate learners' perceived preferences with their actual linguistic performance. However, the idea of individual differences stemming from learner orientations has been with us since the very beginnings of SLA, although there has been conflicting positions on what the consequences of such orientations are in terms of language acquisition. For example, some researchers implicitly have viewed the communication and meaning orientation as facilitative of learning (e.g., Wong Fillmore's 1979 communicators) and the accuracy orientation as debilitating (e.g., Meisel et al.'s 1981 segregative learners). An added complexity in designing future research on learner orientation and task-based performance is that the predisposition towards prioritizing communication or accuracy has been posited to be both an individual difference and a developmentally constrained orientation (cf. Skehan 2002). The topic of learner orientation has nevertheless found renewed interest in recent

research (e.g., Ranta 2002), and the foundations for a research program that fully integrates individual differences with SLA theory has been laid formally by Skehan (2002). Thus, the time seems ripe for future planning studies that systematically address the issue of individual differences.

Language expertise differences: Retrieval vs. rehearsal and self-monitoring

As mentioned earlier, some speakers noted a ceiling in what planning time could buy them, and in several instances a lack of transfer of the plan to the on-line performance was also reported (see Table 7). I speculate that these limitations to the benefits of planning are likely to be related to language expertise, and indeed the strategy and content analyses of the interviews uncovered a few additional differences between the two groups' utilization of planning time. I will focus here on two such areas because of their relevance to theoretical explanations of the benefits of planning.

First, language expertise moderated how learners balanced retrieval and rehearsal operations. For example, the results in Table 4 suggest that more advanced level speakers were able to more equally distribute their efforts to *writing for retrieval* (44%) and *writing for later recall* (31%) and half of them also optimally utilized *rehearsal* strategies during pre-task planning (53%). The reverse pattern was true of low-intermediate speakers. In this group, there was a preference for *writing out for retrieval* (33%) over *writing down for better recall* (no low-intermediate level speaker reported this strategy), and the number of low-intermediate level speakers who reported any *rehearsal* strategies was small (33%). The content analysis of the interviews supports and amplifies this difference.

Among low-intermediate level speakers, there was disappointment and frustration when they described their intense and stressful attempts at retrieving very simple vocabulary during the planning time:

[during planning] I was like, 'Oh no, how come I don't know this word, how come I don't know this word?' I was going 'Think back to way back when you learned the real simple words.' And I was, like, going 'Oh no!' [95011]

Second to retrieval of vocabulary was a salient concern with retrieval of verbal morphology. This was an oft-mentioned area of conscious effort by low-intermediate learners, supported by the opportunity to write out notes. When asked to compare his performance on the planned versus unplanned story, this low-intermediate level speaker nicely summarized this prevalence of cognitive investment in retrieval of vocabulary and verbal forms:

At least there [in the planned narrative] I could organize my thoughts a little bit more. I had time to like remember all the vocabulary to say and all the verbs. [95007]

The picture is different among the advanced level speakers. Resonating with the concerns of low-intermediate learners, many advanced level learners also reported that writing allowed them to engage in lexical searches and in the retrieval of learned verbal morphology. Unlike low-intermediate speakers, however, rehearsal cycles during the preparation stage of the planned task were described in great detail by many advanced level speakers. A representative case of complete and repeated rehearsal of the story is found in this advanced level speaker recount:

I wrote my notes and then read it through, and mentally practiced so that I was actually telling the story for the second or third time as I was speaking for the first time. That gave me the chance to reach, you know, out to the side and point out to more things as I went along. [99024]

Other advanced level speakers reported that they did not attempt to rehearse the complete story in any particular order, but directly concentrated instead on partial rehearsal of problem spots, mainly vocabulary items or phrases:

After writing the notes, I didn't worry about all the other pictures, I kind of ignored them. I just went to these two parts, where I knew I was gonna have problems, and I rehearsed by repeating it. And then I rehearsed the underlined words that I had remembered 'cause I didn't want to forget to use them. I reviewed basically what I'd done. [99032]

Reportedly, then, during pre-task planning many advanced level speakers used a combination of cognitive strategies in particular sequences in order to complete the full planning cycle of 'organizing thoughts', 'retrieving appropriate words and rules', and finally 'rehearsing for immediate performance'.

A second qualitative difference related to language expertise pertains to self-monitoring strategies. The low-intermediate level speakers reported general efforts at being correct during on-line performance (*production monitoring* was mentioned by 75% of them), and some were concerned with assessing the impact of their performance on the listener during their delivery of the narratives (*monitoring the impact of one's performance on the listener* was reported by 33% low-intermediate level speakers). However, they seemed unable to use more concrete tactics for self-monitoring while they prepared for the story. Given the preponderance of retrieval concerns that were reported, it seems fair to say that for this group pre-task planning was mostly consumed with attempts at retrieving rather than monitoring language. By contrast, most self-

Table 8. Self-monitoring strategies during pre-task planning and on-line performance

Strategy	Illustration
Production monitoring	[I: could you hear if you were stopping or correcting yourself?] uh, yeah, it's like, it was like after I say the word is that the right way to say it? I ask myself, so I think maybe a couple of times I changed my tense, like plural [...] cause I'm not really trying to correct myself at the time, I'm just saying what comes into my mind, you know? but then a few times I'm like I ask myself, is it right? [99008]
Monitoring impact on listener	I guess I was trying to see the expression on her face, it's like if she's confused or like- but she looked relaxed, so I said, I guess I'm doing fine [laughs] I'm trying to see if there's confusion in her face it's like, she looked relaxed, so I guess it was okay. [99008]
Auditory monitoring	[I: and what about your articles and your prepositions and all the things you were talking about before?] oh [laughs] I noticed that I always messed them up, like, since I can't write them and I don't see them so I have to, so what I do is I always just say it and then I hear it that it's wrong, so then I say it again and then like I- then they don't match, like you know, the number of persons, the articles, oh yeah then, then I have to I just keep saying them until I figure out which one is the right one and, usually the last one is the right one, I think it's the right one [laughs] [99006]
Visual monitoring	well at first I was kind of pressed for time so I just kind of tried more to just write it out and then I went back and changed, you know, like instead of <i>son</i> {ser copula 'be'} I realized oh no that it isn't <i>son</i> , it's <i>estan</i> {estar copula 'be'} [99003]
Cross-language monitoring	[I: can you give me an example of trying to remember your grammar?] Oh, when to use subjunctive, you know, when they tell him to leave you know to say <i>vayan</i> {subjunctive imperative} instead of something else, or to put <i>le</i> {him, preverbal clitic pronoun} in front of <i>dice</i> {tells} and all those things that I – you don't do in English, that you have to start to think, they don't come automatically [99005]
Style monitoring	I hear myself too you know, I hear myself saying it very slow and I can't get the words, and then it's frustrating [99003]
Double-check monitoring	I was mouthing different possibilities for phrases that I didn't know how to say, and I wrote out what I was saying, and trying to remember also the story in English. Then I looked at everything to see if it fits correctly. And then I guess I was telling it to myself in Spanish to see how it sounds, I reread it a couple of times and rehearsed it. [99016]

monitoring strategies were reported by advanced level speakers, including three monitoring categories that were never mentioned by low-intermediate speakers: *auditory monitoring*, *cross-language monitoring*, and *double-check monitoring* (cf. Table 4). Table 8 provides illustrations for each self-monitoring strategy documented in the advance level learner interviews. In sum, monitoring strategies seemed to be more characteristic of advanced than low-intermediate level speakers.

These differences in the utilization of planning according to language expertise may help explain some of the discrepancies observed in the linguistic outcome findings in the two studies (see Table 3). Specifically, the fuller engagement with self-monitoring strategies by advanced level learners and their balanced commitment of effort to retrieval and rehearsal may help explain the advantage in accuracy, at least for one of the two structures investigated, in the 1999 study. Similar benefits for accuracy with higher proficiency learners are also reported by Wigglesworth (1997b). On the other hand, the prevalence of retrieval strategies among low-intermediate learners, many of them committed to solving lexical problems, is also consistent with the positive finding for lexical complexity as measured by type-token ratio in the 1995 study. This interpretation is consistent with Skehan and Foster's (2001) suggestion that effort committed during pre-task planning to rehearsal is more likely to benefit on-line accuracy, whereas effort invested in retrieval during pre-task planning can be expected to favor on-line complexity.

In the remaining presentation of the results, I will address various findings that revolve around the issue of whether speakers focused on meaning or form. First I present strong evidence that one task feature, the presence of an authentic listener, had important and largely divergent consequences on speakers' conscious allocation of effort towards meaning and form. Next, I examine clear cases of an explicit focus on form by many learners during planning. I finish the presentation of results with a summary of the main findings.

Listener sensitivity and prioritization of communication

A feature of the experimental tasks emerged in the interviews as having shaped speakers' interpretation of task goals: the presence of an authentic listener (see Brown 1995; Brown & Yule 1983). The element of an authentic listener pushed to the foreground the affective and social dimension of task performance in the interviews. For instance, many speakers felt safe narrating for someone who was a peer from the same class, someone who was construed as a listener willing to share some of the responsibility for making meaning in the task (cf.

Lindemann 2002). In fact, the idea of talking to another non-native speaker seemed to relieve some of the tension of the experiment for some learners:

And I'm also relying somewhat on the fact that she's an English speaker [...] maybe if I say something wrong Spanish people would get more confused, I don't know, but I think it'd have been fine if I'd said that for an English speaker. [99011]

Thinking of the listener's needs, however, also created a strong sense of accountability and some stress for other speakers:

I was worried about the presentation, I was worried about her being able to understand what I was trying to say, cause that's the whole point of what we're doing, so that she can say, 'Well, I understand.' And if I couldn't present it in that way then, we're missing the whole point of the exercise. That's where the stress is. She was depending on me. [95008]

Most speakers in both studies interpreted the task as one of conveying the received information in a way that enabled the listener to choose the right pictures and to establish the correct order of events in order to narrate the story accurately in writing (this was indeed the post-task assignment the listener completed while the speaker was being interviewed). This interpretation of task goals appears to have motivated several listener-related strategies. This is the case of *empathizing with the listener*, the only high-frequency social/affective strategy, and *monitoring the impact of performance on the listener* during on-line delivery of the narratives (see Table 4).

In addition, speakers' sensitivity towards the listener impacted upon other strategic choices which indirectly may have had unpredicted consequences for the linguistic products of planning. Namely, these choices can be viewed as having either debilitating or facilitative potential for language development, insofar as they fostered learners' attentional focus on meaning or form. They are summarized and illustrated in Table 9.

Many speakers reported that they expended considerable effort during pre-task planning in organizing the content of the narrative in listener-sensitive ways. For instance, speakers looked for distinguishing features to describe in each picture or they decided to organize the story by picture frame and to retell it in a "segmented" way (see verbatim illustrations in Table 9). In some cases this may have resulted in planned narratives that were rhetorically or propositionally simplified.

Considering the needs of the listener also led to the preference for *approximation* (cf. Table 4) over other compensatory communication strategies when dealing with vocabulary problems, perhaps deterring these speakers from

Table 9. Listener sensitive choices as sources of potential influence on the linguistic outcomes of planning

Choice	Illustration
Organizing the content of the narrative in listener-sensitive ways	<p>I tried to write down something for each picture, something that would distinguish that from all the rest. [99015]</p> <p>I was thinking that she has only a certain amount of pictures, and if I keep it segmented then she might be able to easier identify what I'm talking about. [95011]</p> <p>I wanted to be able to say what was going on in each picture [...] so it was easier for her to choose exactly which one she wanted – I did it that way." [95014]</p> <p>[after third story, unplanned] I was trying to figure out how to differentiate between all these pictures, for him to pick up the right ones, trying to think of something that could identify the pictures. [99026]</p>
Coming up with simple vocabulary or approximation to it good enough to ensure listener comprehension of the story	<p>[for unknown words] I would use something else instead, just mainly so that she could understand what I was saying, not necessarily to get the exact right verb for that, but you can actually know what was going on, so that she kind of had an idea of what it was. [99009]</p> <p>I tried to pick up easy words that he would understand what I was trying to describe. [95007]</p> <p>I tried to make it very clear and simple so that she can get the point across, she can know even if the vocabulary was wrong, at least she can look at the picture and kinda get the idea of what I was trying to say. [95008]</p>
Focusing away from "language", keeping it simple	<p>with the three exercises I was very conscious of trying to make it very simple and very clear, and I'm not very worried about the speed or the correctness of the grammar [99011]</p>
Reluctance to self-correct on-line, pressure to keep going	<p>Then I just said <i>children</i> {kids}, <i>niñas</i> {girls}, that's it, you know, I couldn't think of the word- but I think one of them is a boy, and I just noticed that, and I said "Forget it, just keep it a girl because she's gonna get confused," so I just kept it a girl [99009]</p> <p>When I hear a mistake I know it's a mistake, but I couldn't stop and fix it, because then that messes up the person I'm telling, so I'm like "Wait, I have to focus on the person I'm telling the story, not whether I got my mistake." [99028]</p> <p>I felt that if I corrected myself then she may get too confused [...] so I just kind of kept going and let it slide. [95006]</p>
Slowing down during performance for the listener	<p>Some of it [the time spent pausing] I was thinking, and some of it I was giving my partner time to get the pictures. [95007]</p>

Table 9. (continued)

Choice	Illustration
Thinking of grammar that is essential for listener's understanding	Cause there were two of them [two people in the story], so I had to make it in the <i>ellos</i> {they} form, I don't know. [...] ...'Cause I think that that would have helped her cause she would know how many people were there in the picture and stuff. [I: So conjugating the verb was important for the meaning?] Yeah, I guess so. [95012] [in the third story, planned] I didn't care anymore about making mistakes, but then not like mistakes where you use the wrong verb tense or the totally wrong word and you throw the person off. [99009]

pushing their lexical retrieval efforts to their limits. Furthermore, as the illustrations in Table 9 make clear, several speakers may have sacrificed propositional, lexical, or syntactic complexity for the sake of the listener. Some learners stated they tried to be consistent and simple in the way they told the story to avoid confusing their partner. Others went as far as to characterize their approach to the experimental tasks almost entirely from the perspective of the listener, and they stated that this led them not only to avoid lexical or grammatical complexity but also to focus away from a concern with language, be it worrying about being fluent or thinking about grammatical accuracy (cf. verbatim illustrations in Table 9). Finally, there were also reports of reluctance to self-correct during online delivery of the story for fear of confusing the listener. More specifically, several speakers commented on the fact that they avoided stopping and felt a pressure “to keep going” and let utterances in need of self-repair “slide” in order to keep the listener on track. In addition, however, fluency may have been affected by the speakers' orientation towards the listener in other ways as well. When this speaker was asked why he paused sometimes during the story, he responded:

Some of it [the time spent pausing] I was thinking, and some of it I was giving my partner time to get the pictures. [95007]

The choices discussed so far point at potentially debilitating effects for language development, since they motivated a concern with meaning and communication that is in conflict with the heightened focus on form that the planning literature posits as beneficial for language development (e.g., Skehan & Foster 2001). In one respect at least, however, the listener-oriented interpretation of the task also had a beneficial impact on learners' attention to language. This is shown in the last theme illustrated in Table 9. Namely, the concern with the listener and with getting the message across to their partner led several speak-

ers to attend to aspects of grammar that were perceived as essential for the listener's understanding, in what can be described as a heightened process of meaning-form mapping. This speaker, for instance, paid attention to conjugating verbs during the planning time for the sake of referential clarity and comprehensibility:

I think I was trying to worry about conjugating the verbs in the correct form, and since there are two main characters in the story trying to make sure that I did differentiate between 'they' when, when both of them were together and when they were, you know, singular. [95008]

To summarize, a prevalent listener-sensitive interpretation of the task guided speakers' approach towards planning and performance and may have had varying potentially debilitating and facilitative consequences for learners' attention to the language code: (a) It may have deterred some learners from engaging in propositional, lexical, and/or syntactic complexity, because they prioritized meaning and communicative efficiency (Skehan & Foster 2001) for the sake of ensuring listener's comprehension; and (b) it may have made some speakers avoid self-corrections during online performance and pressured them to prioritize fluency over accuracy (cf. Yuan & Ellis 2003) for the sake of preempting listener confusion. At the same time, however, (c) it may have pushed some speakers to make form-meaning connections that were perceived as essential in the task at hand (Loschky & Bley-Vroman 1993) so that the listener could make sense of the story (e.g., subject-verb agreement in Table 9).

Explicit focus on form during planning

There is also additional evidence in the interviews of more explicit kinds of attention to formal aspects of the language during planning. This was observed in widely documented remarks about language issues, when speakers reported a conscious effort to check, verify, and correct their speech by drawing on explicit kinds of knowledge they had about their first language, the target language, and about themselves as language learners. These cases are listed and illustrated in Table 10. The use of these funds of explicit knowledge to self-regulate L2 performance suggests that for some learners pre-task planning facilitated a heightened awareness about one's own linguistic resources vis-à-vis performance demands, a form of noticing the gap and noticing holes (Swain 2000).

Two strategies indicate that learners utilized knowledge of their first language as a form-focusing resource: analytical monitoring and translation. An-

Table 10. Speakers' attention to form drawing from explicit kinds of knowledge

Category	Example
Crosslanguage monitoring	[Interviewer: you also crossed out things in your paper. What were you thinking and why did you cross out things?] Because there is sometimes when I would try to say something but... I was thinking in English too much and it doesn't come out in Spanish like that, so when I can see it then I think <i>No that's not the way it goes</i> so then I try another way to say it in Spanish and that's what I need to stick with. [99029]
Translation	I went over what the story was about in English and then I thought I thought my response in English and I tried to translate it into Spanish [95008] basically I think that I would look at the picture and say what I thought it was in English and then write down what I wanted to... and how I would say that in Spanish. And if I couldn't say it in Spanish, I do another way to say it in English so that I could say it in Spanish [95014] thinking back over it, what I had in Spanish, thinking back over it in English, what I was actually saying on paper [95006]
Using metacognitive knowledge of own as learner, for instance for prognostic evaluation	I noticed that if I can write it out then I come up with a whole lot of words but then if I just have to tell and just say it right, then and there, then I limit myself to small vocabulary, like the vocabulary I learned from my first year of Spanish, that's my strongest, so I noticed that. [99006] that's how I remember things, by writing them [...] and I always have a hard time with the conjugations [99002] I'm never quite sure about the endings, where, you know, masculine feminine endings of things I can usually remember the root of the word but always lose it on the endings and get the <i>el</i> {the-masculine} and the <i>la</i> {the=feminine} usually mixed up [99005]
Using metalinguistic knowledge	Like if the verb was <i>invitar</i> I'd go, <i>-o -as -a ... invita</i> [I: So you would actually go through the verbs?] Yeah. I do that all the time. And I want to make sure that I got at least the right person." (+95011) [I: Did you have any time to think of the grammar also?] Yeah. I had some time to think of it. Like that personal <i>a</i> thing {rule for prepositional marking of human direct object} and, you know, the possessive adjectives and, the verb tenses, saying them right, and gender agreement, things like that. And like <i>ser</i> and <i>estar</i> , when to use those... Yeah, I was thinking of that as well as trying to say it. [95006]
Attending to grammar of low communicative value	Reflexive verbs: The grammar and stuff too, I don't know how to put it, like <i>se preocupa mucho</i> {s/he worries a lot}? "it worries," I know it's supposed to mean "it worries them" so I guess it is <i>le preocupa</i> {he worries}, but I get that confused a lot, like even if I wanted to say "the boys hit the ball," would it be <i>se pegan</i> {they hit each other} or would it just be <i>pegan</i> {they hit} [99003]

Table 10. (continued)

Category	Example
	<p>Articles: [I: What did you do during the preparation time?] I wrote all the story [laughs] cause then I can see like where I was supposed to put articles and that stuff so if I can see it then it's a little easier for me, so that was better [99006]</p> <p>Subjunctive: When I said "están contentos que hayan..." I remembered that you have to use subjuntivo for some kind of opinion or emotion [I: Did you remember that when you were writing?] Right, yeah, when I was writing it's like I remembered it, so I used it. [99015]</p>

alytical monitoring refers to a conscious comparison of the expression of an idea in the L1 and the L2 in order to judge the 'transferability' of an L1 utterance or the 'Spanishness' of an L2 utterance. Very few speakers (four out of the 44, all of them from the advanced level group) reported the use of such a comparison. Translating, on the other hand, was widely reported among the L2 Spanish speakers. The fact that the task started with an audio-recorded version of the story in the L1 may have biased the participants towards translating more than normal. Nevertheless, the ample documentation of translation during planning adds on to the robust finding that language learners use the L1 as a cognitive tool during task performance (see review in Swain & Lapkin 2000; and for monologic types of language performance, see Cohen et al. 1998).⁵

Speakers also made ample use of their explicit knowledge of themselves as language learners (or what other researchers have called metacognitive knowledge, e.g., Wenden 2001). Most speakers were able to provide evaluative comments about where they "have most trouble" or where they always "mess up," offering remarks about what vocabulary and grammar areas were "their weakest." A specific application of such knowledge to strategic behavior was *prognostic evaluation*, reported by a few learners in both groups (cf. Table 4 and illustrations in Table 10).

Certainly, these learners displayed a good knowledge of the pedagogical grammar instructed in most Spanish lower-division university courses. This speaker, for instance, was clearly capable of reciting most pedagogical grammar rules of Spanish and claimed that she attended to all those aspects during the tasks:

[I: Did you have any time to think of the grammar this time also?] Yeah. I had some time to think of it. Like that personal *a* thing {rule for prepositional marking of human direct object} and, you know, the possessive adjectives and, the verb tenses, saying them right, and gender agreement, things like that. And

like *ser* and *estar* {the two Spanish copulas}, when to use those... Yeah, I was thinking of that as well as trying to say it. [95006]

Indeed, a considerable attention to form was documented not only with task- or listener-essential grammar, as discussed in the previous section, but with morphology that was less meaningful or, in VanPatten's (2002) terms, of lesser communicative value, such as reflexive pronouns, articles, and the subjunctive (all illustrated in Table 10).

The heightened attention to morphosyntax that pervaded the interviews (see Tables 6, 9, and 10 for multiple illustrations) is worth commenting on because it runs counter to the primacy of attention to lexis and meaning that has been typically found in task-based performance by ESL learners (e.g., Pica 1994). I have argued elsewhere (Ortega in press) that this may be typical of L2 learners functioning in foreign language contexts. Findings reported across several recent task-based studies with foreign language learners offer support for this hypothesis (see Buckwalter 2001; Iwashita 2001; Mackey et al. 2000; Pelletieri 2000).⁶ Exploring possible reasons, Iwashita (2001:278) suggests that foreign language learners are more aware of and possibly more resourceful with negotiating morphosyntax of the target language during pedagogical practice because much early FL instruction places a good deal of emphasis on the explicit learning of morphosyntax, possibly deemphasizing vocabulary to some extent. Buckwalter (2001:392) further speculates that by their very nature so-called strong-morphology languages, such as Spanish, Japanese, or Italian, require of learners much more attention to morphosyntax than weak-morphology languages like English. The claim that foreign language learners are generally more attuned to morphosyntax than ESL learners, in turn, means that we must be cautious not to overgeneralize the finding of explicit focus on form during planning in the present interviews to non-foreign language learner populations.

Summary of interview findings

To summarize, the strategy and content analyses of the speaker interviews offered the following findings. First and foremost, the types of strategies documented and the benefits identified by speakers in the two studies point at the centrality of retrieval and rehearsal operations during pre-task planning. Retrieval operations enabled by extra time, in particular, were pervasive and are consistent with the gains in syntactic complexity observed in both studies (cf. Table 3). In addition, learners' strategic attention to form was clearly doc-

umented in retrieval, monitoring, and rehearsal activities directed to specific grammatical forms, which were greatly supported by the opportunity to write notes (cf. Tables 6 and 8). Second, I have argued that individual differences and differences in language expertise moderated learners' perceptions of planning and, possibly, their ability to benefit from it. Individual differences in speakers' own preferences and perceptions of what learning and using an L2 entails may have guided their efforts during pre-task planning to what they viewed as important, namely communication or accuracy. Language expertise, on the other hand, seems to have filtered the nature of benefits afforded by planning. Among advanced level speakers it allowed a more balanced commitment of effort to retrieval and rehearsal and a fuller engagement with self-monitoring strategies, while among low-intermediate speakers it fostered retrieval strategies committed to solving lexical and verbal morphology problems. This interpretation is consistent with the differences in findings regarding accuracy and lexical complexity observed between the two studies (cf. Table 3). Third, the presence of an authentic listener was a task feature that encouraged many learners to orient to the listener's needs and to prioritize getting the message across to the listener over being accurate, fluent, or complex. This listener orientation may have deterred some learners from engaging in propositional, lexical, and/or syntactic complexity and it may have made some speakers avoid self-corrections during online performance and pressured them to prioritize fluency over accuracy. At the same time, it also led to a heightened process of meaning-form mapping, by priming some learners to attend to certain aspects of grammar that were perceived as essential for the listener's understanding. Finally, it was found that many learners also paid attention to formal aspects of the language of low communicative value and exploited different funds of explicit knowledge to guide their strategic behavior during planning.

Conclusion

In the end, then, what do I conclude from my analysis of the interview data? What do learners do when they plan, and what do they attend to? In the most metalinguistic sense, if we take Swain's (2000) concept of negotiation of form, then in spite of the narrative task in the two studies being monologic rather than collaborative, pre-task planning created the mental space for learners to negotiate with themselves many aspects of the language and allowed them to utilize various funds of explicit knowledge that guided their conscious attention towards areas in which they were well aware of holes and gaps vis-

à-vis the specific task demands. If we take attention to form as synonymous with a concern for being accurate and/or being sophisticated while using the L2, as Skehan and Foster (2001) do, then learners like [99005] and [99015] were undoubtedly concerned with being accurate during pre-task planning, and many other learners reported paying attention to elaborating details of the content and the language in the context of pre-task planning. There were, however, many occasions in which most speakers seemed to be driven by the “meaning-first” principle (and its task-specific counterpart in the two studies, “the listener-first principle”) that Foster and Skehan consider debilitating for language learning.

Ultimately, I would like to suggest the analysis of the interviews presented in this chapter challenges this kind of dichotomy of “attention to form” versus “attention to meaning,” because during meaningful second language production (and when preparing for it) learners engaged in solving form-in-meaning problems. Thus, the interviews give us insight into the most psycholinguistic sense of focus on form as defined by Doughty (2001), Long and Robinson (1998) and Robinson (2001a). The retrospective data document clear and frequent instances in which learners seemed to pay attention to the inextricable relationship between form and meaning, simultaneously holding in long-term memory considerations regarding the message to be conveyed and the essential formal resources to convey it. In this sense, task-based performance for these learners in these two studies “ma[de] functional co-ordinates of grammatical structure (as well as speech acts and lexis) available to learners, along with their purely formal aspects” (Robinson 2001a:292). That is, the opportunity to plan prior to the task afforded speakers the time and space to self-regulate their performance, to weigh and orient themselves to the task demands, and to direct their resources and attention to the language needed to complete the task. In spite of holding a meaning-oriented interpretation of the task (or perhaps more precisely because of it), learners paid attention to form during planning without any specific instructions to do so.

At a minimum I believe we need to begin thinking of a new metaphor, in which there is a debilitating kind of attention to meaning (the one discussed by Skehan & Foster 2001) and a facilitative kind of attention to meaning, the one supported by functionalist theories of language learning (N. Ellis 2002; MacWhinney 2001) and by the focus on form position (Doughty 2001; Long & Robinson 1998; Robinson 2001a). In my view, the disagreement is one of two (closely related) positions: focus-on-forms versus focus-on-form. The focus-on-forms position (Skehan and Foster 2001; VanPatten 2002), drawing on limited capacity theories of attention, emphasizes the dichotomization of form

and meaning. The focus-on-form position (Doughty 2001; Long & Robinson 1998; Robinson 2001a), drawing on both symbolic and emergentist functional theories of second language learning, disagrees with the form-meaning dichotomy metaphor and even suggests it may be an impossible one from a theoretical standpoint (see Robinson 2003). This is the joint position of DeKeyser, Salaberry, Robinson, and Harrington (2002), who contend that “simultaneous attention to form and content is clearly possible” (p. 809) in adult second language processing. Ironically, VanPatten’s (2002) Input Processing Instruction is crucially based on the assumption that learners will best learn a new form when they are forced to pay attention to certain formal properties *while processing it meaningfully*. Thus, in the end we may all be talking about the same form-in-meaning qualities of optimal language processing and language learning at some basic level, and an open discussion of our ‘form’ and ‘meaning’ metaphors may be a fruitful step towards coordinating and strengthening research programs for planning and task-based language learning.

The interview findings, triangulated with the linguistic outcomes in both studies, support the interpretation that planning “create[d] a space for the learner to assess task demands and available linguistic resources and to prioritize strategic allocation of effort and attention accordingly” (Ortega 1999: 138). This is indeed the kind of beneficial impact on long-term acquisition that I would like to postulate for pre-task planning: It enables a conscious shift to a learner-driven and learner-regulated focus on form, or put in simpler terms, it fosters learner’s attention to language as a meaning-making tool.

I would like to conclude with a reflection on the promise of the process-product approach proposed here for the investigation of task-based L2 performance. The analysis of the interviews allowed me to uncover a number of issues that I would have missed by analysis of the linguistic product alone. More globally, the analyses presented in this chapter force us to ponder on the extent to which affective and social elements of communication can shape learners’ cognitive and linguistic behavior, even in a monologic task under restrictive experimental conditions. This underscores the need to expand the present research focus on task requirements and linguistic outcomes and to consider social and affective dimensions of task performance in future research. For over a decade now, Yule and his colleagues (Yule 1996; Yule & Powers 1994; Yule, Powers, & Macdonald 1992) have contributed insights in this area from a discourse analysis perspective. A similar concern has inspired recent investigations from either psychological (e.g., Dörnyei & Kormos 2000; Dörnyei 2002; Robinson 2001b, on task engagement) or sociocultural perspectives (e.g., Platt & Brooks 2002, on task engagement; Morris & Tarone 2003; Storch 2002, on

interpersonal dynamics). As these different strands of research show, it will be both feasible and profitable to address affective and social dimensions of task-based performance from a variety of theoretical traditions in future research. Whether through discourse analysis, psychological questionnaires, or post-task interviews, a process-product approach that considers the full landscape of variables contributed by task, learner, and linguistic outcomes, has the potential to illuminate our theories of the role of planning in task-based language learning.

Notes

1. In this and all other excerpts from the interviews, participant comments are verbatim and any additional information needed to make the meaning clear is given in squared brackets. Spanish words are italicized and translated or explained in curly brackets. At the end of each quote, the speaker is identified with a unique number, which starts with 99 or 95 to signal the study in which the speaker participated.
2. The experimental procedures were different in the two studies in two respects. First, in keeping with most previous research, the 1999 study speakers were given 10 minutes, whereas in the 1995 study speakers were given eight minutes to plan because during piloting with similar-level students it was observed that speakers could not make use of the full 10 minutes. Second, in the 1995 study, the low-intermediate level speakers did two narratives each, the first one with preparation and the second one extemporaneously. This was done because of logistical constraints to keep the sessions under an hour. By contrast, in the 1999 study the advanced level speakers did three narratives, of which one was planned. Unbeknownst to the 1999 participants, the first narrative task served as a familiarization phase prior to the experimental tasks (although the post-task interview for this familiarization narrative was also recorded and analyzed), whereas the second and third tasks were the focus of the study. The planning condition was counterbalanced randomly across the 32 speakers: half were asked to plan for the second narrative and to do the third extemporaneously, and the other half planned the third narrative and did the second one extemporaneously.
3. I am most grateful to Megan Thompson for her help collecting the data in the 1995 study and transcribing and coding those initial interviews. Megan was trained in qualitative research methods and had no investment in task-based language learning theories of planning, both useful qualities that greatly contributed to our joint analysis of the 1995 interview data.
4. In addition, fewer cognitive and social/affective strategies may have been reported because instructions in the two planning studies explicitly prohibited the use of several such strategies. The strategies explicitly banned were: switching to the L1, using mime, selecting the topic, resourcing (e.g., using a dictionary), questioning for clarification, and cooperating with others.
5. It is interesting to note that there were also nine speakers (eight in the advanced group and one in the low-intermediate group) who viewed translation as a counterproductive strat-

egy and advocated the maxim of 'thinking in the target language as much as possible'. In their study of speaking and learning strategy training, Cohen et al. (1998) observed similarly divided attitudes towards translation among their learners.

6. Buckwalter (2001) found a balance of morphosyntactic and lexical repairs in frequent self-corrections made by first- and second-year Spanish learners during task-based interactions. Similarly, Pellettieri (2000) observed evidence of self-correction involving morphosyntax when she inspected her Spanish learners' use of the backspacing key while doing a computer-mediated collaborative task. Iwashita (2001) found more syntactic than lexical modifications in task-based interactions by fourth-semester Japanese learners. More tenuously, Mackey et al. (2000) noted that their L2 Italian learners were somewhat better (although still far from ideal levels) at identifying negative feedback on morphosyntactic errors by comparison to their ESL learners.

CHAPTER 4

The effects of focusing on meaning and form in strategic planning

Jiraporn Sangarun

Suranaree University of Technology, Thailand

Introduction

As Ellis pointed out in Chapter 1, there is now substantial evidence that providing learners with the opportunity to plan strategically can have beneficial effects on task-based production and may also assist second language (L2) development. It can also be noted that strategic planning has increasingly been integrated into communicative activities (Brumfit 1984; DeKeyser 1993; Morrow 1979), task-based instruction, and task design (Long & Crookes 1993; Nunan 1991; Skehan 1996a). However, while strategic planning is now recognized as an important factor in L2 production how it assists production has not been fully explained. In particular, there has been little research that has examined what learners do while they plan (but see the chapter by Ortega) and not much more that has attempted to examine the effect of manipulating learners' attention onto different aspects of production while they are planning. Many studies (Wigglesworth 1997; Ortega 1999) have given no specific instructions to learners as to how they should plan, allowing them to attend to form or meaning or both as they choose. However a number of studies have examined the effects of the focus of strategic planning – whether on form or meaning – on learners' production. In some studies (Ellis 1987; Foster & Skehan 1999; Williams 1992) learners have been guided to pre-plan *either* meaning *or* form, while in others (Crookes 1989; Foster & Skehan 1996; Mehnert 1998; Wendel 1997) they have been guided to pre-plan *both* meaning and form.

Previous research on the focus of strategic planning is limited and inconclusive. Three main gaps are evident. First, none of the previous studies investigated the effects of the focus of strategic planning upon participants'

cognitive processes. Consequently, it is unclear whether the participants focused on meaning, on form, or on both meaning and form during strategic planning in accordance with the instructions they had been given. Second, most of the studies failed to investigate the effects of the focus of strategic planning on the participants' actual application of their plans. Finally, rather than comparing the effects of different kinds of foci of strategic planning on speech performance, the comparison made was between a general planning condition and a no-planning condition.

This chapter reports on a study that was designed to shed light on the gaps in the previous research. It has the following objectives: (1) to identify and describe in detail the strategic planning processes that resulted from a focus on (a) meaning, (b) on form, and (c) on meaning and form; (2) to explain the effects of these three foci of strategic planning on the learners' application of their plans; and (3) to examine the effects of these three foci of strategic planning on the quality of speech.

Effects of the focus of strategic planning on speech performance

Foster and Skehan (1999) investigated the effects of meaning-focused strategic planning and form-focused strategic planning on oral production. They found that these two foci did not produce different effects on the accuracy, complexity, and fluency of speech.

Four studies addressed the issue of meaning/form-focused strategic planning. Crookes (1989) and Mehnert (1998) guided their participants to plan both the meaning and the form of their speech. Crookes found that the meaning/form-focused strategic planning condition resulted in significantly higher complexity but not in higher accuracy than the minimal strategic planning condition. Mehnert reported that, under the meaning/form-focused strategic planning condition, speech manifested significantly higher accuracy and fluency and a broader lexical range than speech under the minimal strategic planning condition. Foster and Skehan (1996) compared the effects of (1) meaning/form-focused strategic planning, (2) undetailed strategic planning and (3) minimal strategic planning on EFL learners' speech in three experimental tasks: personal information exchange, oral narrative, and decision making. They found that, under the meaning/form-focused strategic planning, speech was more complex and fluent (for all three tasks), and more accurate (for two of the three tasks) than speech under the minimal strategic planning condition. In addition, they found that the meaning/form-focused strategic

planning condition promoted significantly higher speech complexity (for all three tasks) and higher fluency (for one of the three tasks) than the undetailed strategic planning condition. However, the two strategic planning conditions (i.e., meaning/form-focused and undetailed) did not have a differential effect on accuracy. Wendel (1997) guided his participants to plan the meaning, vocabulary, and discourse structures of their narratives. He found that this meaning/form-focused strategic planning, when compared to the minimal strategic planning condition, led to greater complexity and fluency but not greater accuracy or lexical variety. Lastly, Wigglesworth (1997) found that a one-minute unguided strategic planning condition led to significantly higher accuracy, complexity, and fluency in the case of cognitively demanding tasks but only for high proficiency participants. Low proficiency participants did not manifest any advantage from the opportunity to plan. She hypothesized that the higher proficiency participants probably planned both meaning and form, while the lower proficiency participants probably planned only meaning.

In summary, the research cited above suggests that (1) meaning/form-focused strategic planning is more effective than the minimal strategic planning in promoting accuracy, complexity, and fluency (2) the three foci of strategic planning (i.e., meaning-focus, form-focus, and meaning/form-focus) seem to have similar effects in promoting the accuracy, complexity, and fluency of speech. However, there is a clear need to investigate whether the specific foci of strategic planning have differential effects on task-based performance.

Method

Operationalization of strategic planning conditions

The four strategic planning conditions that were used in this study were operationalized as follows:

1. In minimal strategic planning (MinP) the participants were allocated no time for strategic planning.
2. In meaning-focused strategic planning (MP) the participants were given fifteen-minutes for strategic planning time. In addition, they were instructed in how to plan the meaning of their intended speech.
3. In form-focused strategic planning (FP) the participants were given fifteen-minutes of strategic planning time. In addition, they were instructed on how to plan the form (i.e., vocabulary, transitional words or phrases, and grammar) of their intended speech.

4. In meaning/form-focused strategic planning (MFP) the participants were given fifteen-minutes of strategic planning time. In addition, they were instructed to plan both the meaning and the form of their intended speech.

Research question and hypotheses

The main research question addressed in this study is:

What are the effects of the MP, FP, and MFP on: (a) EFL learners' cognitive strategic planning processes; (b) the way they apply their strategic plans; and (c) the quality (i.e. accuracy, complexity, and fluency) of the speech they subsequently produced?

The following three groups of hypotheses relating to this research question were formulated:

Group A: Hypotheses regarding participants' cognitive strategic planning process

- Hypothesis 1 When the researcher, through written instructions, encourages MP, participants will direct most of their attention towards planning relating to meaning.
- Hypothesis 2 When the researcher, through written instructions, encourages FP, participants will direct most of their attention towards planning relating to form.
- Hypothesis 3 When the researcher, through written instructions, encourages MFP, participants will divide their attention equally between planning relating to meaning and relating to form.
- Hypothesis 4 Participants under the MP condition will direct more attention towards planning relating to meaning than will participants under the MFP condition. In turn, participants under the MFP condition will direct more attention towards planning relating to meaning than will participants under the FP condition.
- Hypothesis 5 Participants under the FP condition will direct more attention towards planning relating to form than will participants under the MFP condition. In turn, participants under the MFP condition will direct more attention towards planning relating to form than will participants under the MP condition.

Group B: Hypotheses concerning participants' application of their strategic plans

- Hypothesis 6 Participants under the MFP condition will use more planned ideas and fewer unplanned ideas in their speech than will participants under the MP condition who, in turn, will use more planned ideas and fewer unplanned ideas than participants under the FP condition.
- Hypothesis 7 Participants under the MFP condition will use more planned grammatical structures and fewer unplanned grammatical structures in their speech than will participants under the FP condition who, in turn, will use more planned grammatical structures and fewer unplanned grammatical structures than participants under the MP condition.

Group C: Hypotheses concerning the quality of speech

- Hypothesis 8 Planned speech will be more accurate, complex and fluent than speech that has been minimally planned.
- Hypothesis 9 When MFP is encouraged, participants will produce speech that is more accurate than when FP is encouraged. In turn, when FP is encouraged, participants will produce speech that is more accurate than when MP is encouraged.
- Hypothesis 10 When MFP is encouraged, participants will produce speech that is more syntactically complex than when MP is encouraged. In turn, when MP is encouraged, participants will produce speech that is more syntactically complex than when FP is encouraged.
- Hypothesis 11 When MFP is encouraged, participants will produce speech that is more fluent than when FP is encouraged. In turn, when FP is encouraged, participants will produce speech that is more fluent than when MP is encouraged.

Research design

A 4×2 research design was employed (see Table 1). The strategic planning condition, the first independent variable, was a between subject factor with four levels: (1) MinP, (2) MP, (3) FP, and (4) MFP. The second independent variable, task type, was a within-subject factor having two levels: an instruction task and an argumentative task.

Table 1. Research design

Group	Number of participants	Strategic planning conditions	Tasks
1	10	Minimal strategic planning (MinP)	Instruction Argumentative
2	10	Meaning-focused strategic planning (MP)	Instruction Argumentative
3	10	Form-focused strategic planning (FP)	Instruction Argumentative
4	10	Meaning/form-focused strategic planning (MFP)	Instruction Argumentative

Participants

The participants in the present study were 40 Thai Grade 11 EFL participants between 16 and 17 years of age in a high school in Thailand. Ten were male and thirty female. They had been studying English from 6 to 12 years, reflecting the Thai National Scheme of Education, in which foreign language learning is optional. They were all at an intermediate level of English proficiency, as indicated by their grades in the two required English courses they had taken. All were studying English at school for six hours per week at the time of the study. To ensure that all participants in the four experimental groups were of a similar level of proficiency prior to the onset of the study, two steps were taken. First, I assigned the participants into ten matched quartets, relying on their grades in their completed English and their length of English study. Second, I randomly assigned each member of these matched quartets to one of the four strategic planning conditions.

Experimental tasks

Both an instruction task and an argumentative task were used in the study. The instruction task, adapted from Mehnert (1998), required the participants to leave a message on a telephone answering machine, telling an English-speaking friend that they could not meet her/him at the train station, as had been earlier agreed, because of an important test scheduled at the same time the English speaking friend was due to arrive. As a result, the friend had to proceed to the callers' school by herself/himself. The argumentative task, on the other hand, required a monologue on the topic of "high school uniforms," in which the par-

ticipants provided their opinions concerning whether or not Thai high school participants should wear uniforms. The instruction task was considered to have lower cognitive and linguistic demands than the argumentative task, according to Skehan's (1996a) three criteria for task grading, i.e., cognitive complexity, code complexity, and communicative stress. Participants carried out the two tasks in order of difficulty, (i.e., the instruction task preceded the argumentative task). This was because previous research indicates that there tends to be a greater effect of task order when a more difficult task precedes a simpler one (Mehnert 1998).

Strategic planning instructions and note-sheets

The strategic planning instructions and note-sheets used in previous research had six main characteristics (Crookes 1989; Ellis 1987; Foster & Skehan 1996; Mehnert 1998; Wendel 1997; Wigglesworth 1997; Williams 1992). First, they guided participants in planning the meaning, discourse structure, vocabulary and/or grammar of their speech. Second, they reminded participants to consider listeners' needs. Third, they advised participants to plan their speech in their second/foreign language. Fourth, they instructed participants not to write down in detail everything they intended to say. Fifth, they instructed participants to make written notes on a piece of paper. Finally, they reminded participants that they could not use their notes as they spoke.

In order to decide on the planning instructions and note-sheets for the present study, I carried out three pilot studies, in which seven participants – four Thai, two Iranians, and one Canadian – volunteered to participate. The Iranian and Canadian participants were included in the pilot studies, so that I could obtain a base-line with which to compare the Thai participants.

The first pilot study had one major objective, namely, to investigate the strategic planning strategies used by EFL and ESL learners. Three participants were involved. Each participant completed an instruction and an argumentative task, as described above. The participants were given 10 minutes to plan and were allowed to do so freely. They were asked to record their plans in English, to refrain from writing down everything they wanted to say in detail, and to make notes on a blank paper. In addition, they were reminded that they could not use their notes when they spoke. This pilot study showed (1) the three participants put their main focus on the planning of meaning, devoting little attention to the planning of form and (2) in general, they did not plan the meaning of their speech effectively.

The second pilot study had two objectives; (1) to investigate whether it was possible to guide participants to redirect their main focus from the planning of meaning to the planning of form, (2) to test instructions, adapted from those used by Foster and Skehan (1996), for guiding participants to plan the form of their speech. Two participants were involved in this second pilot study, again completing the instruction and argumentative tasks. However there were two differences between the first and second pilot studies. First, the participants in the second pilot study were asked to plan only the form and not the meaning of their speech, whereas participants in the first pilot study were allowed to plan freely. Second, they were given a specially designed note-sheet having two sections, one for vocabulary notes and the other for grammatical notes. This second pilot study showed that (1) it was possible to guide participants to redirect their main focus onto form, although it proved impossible to bar participants from thinking about meaning entirely and (2) the planning instructions were not sufficient to guide participants to plan the form of their speech effectively. These instructions only told participants that they must plan the vocabulary and grammar of their speech without specifying how they were to do this.

The third pilot study had two objectives: (1) to test the instructions for guiding participants to plan the meaning of their speech that I had adapted from Foster and Skehan (1996) and (2) to develop more effective instructions for guiding the planning of form. The participants carried out the same instruction and argumentative tasks as in the first and second pilot studies. One difference between the third pilot study and the second was that the participants in the third pilot study were provided with both instructions for the planning of meaning and the planning of form. Results showed that neither of the two sets of instructions were effective, suggesting the need for more detailed instructions on how to plan.

Based on (1) Levelt's (1989) model of speech production (see Chapter 1), (2) the planning instructions and note-sheets used in previous L2 research on strategic planning, and (3) the results obtained from the three pilot studies, I formulated seven principles for developing strategic planning instructions and note-sheets (see Table 2). On the basis of these principles, strategic planning instructions and note-sheets for the MP, FP, and MFP conditions were developed (i.e., MP: Principles 1, 4, 5, 6, and 7; FP: Principles 2, 4, 5, 6, and 7; MFP: Principles 1 to 7 (see Appendices 1 to 4 which show the instructions and note-sheets for the instruction and argumentative tasks under the MFP condition).

Table 2. Principles for developing strategic planning instructions and note-sheets

Principles	Rationale
<p>1. Instructions that aim to guide L2 learners in planning the meaning of their speech should:</p> <p>(a) guide the learners to generate appropriate information for their speech by reminding them of the kind of information needed for the speech;</p> <p>and (b) help the learners shape the information of their speech according to the appropriate discourse structure of the target speech.</p>	<p>Levelt (1989) hypothesized that in the conceptualizing stage, speakers first set their communicative goals, next, retrieve the information needed to express these goals and, finally, select the information that will fulfill their intended goals. For this reason, a set of meaning-focused strategic planning instructions should consist of an instruction that will remind participants of the kinds of information that are needed for each speech.</p> <p>Levelt (1989) hypothesized that after speakers select the information that will achieve their communicative goal, they then order and shape it, using a knowledge store for speech processing that contains both discourse models and situational and encyclopedic knowledge. Hence, to plan meaning effectively, participants need: (1) instructions that will remind them of the appropriate discourse structures for each target speech; (2) instructions that will guide them in how to shape their information according to the discourse structures.</p>
<p>2. Instructions that aim to guide L2 learners in planning the form of their speech should:</p> <p>(a) guide the learners: (i) to plan both vocabulary and grammar for their speech; (ii) to plan vocabulary before grammar.</p>	<p>Levelt (1989) assumes that to convert a preverbal message into an internal speech, speakers must first select language for the preverbal message by retrieving lexical items from their mental lexicon. This action, in turn, will activate the syntactic building procedure that will carry out grammatical encoding.</p>

Table 2. (continued)

Principles	Rationale
(b) guide the learners: (i) to refrain from selecting several words for one meaning; (ii) to plan sufficient vocabulary to cover the entire message they wish to transmit;	This is based on the results of the third pilot study, in which one participant included several words for one meaning while another failed to include sufficient words to cover the entire message she wished to convey. Neither of these practices is effective. In the first case, the participant had to select from several possibilities the word he actually wanted to use as he spoke. In the second case, the participant had to think of and add other words as she spoke.
(c) guide the learners to focus on planning those grammatical structures that are likely to have a significant impact on the target speech;	This is based on the results of studies by Hulstijn and Hulstijn (1984). This study suggests that guiding participants to focus their attention on grammatical structures promotes the correct use of those structures. This assumption also relies on the results of the second and third pilot studies, which showed that all of the participants failed to focus on planning grammatical structures that were likely to have an important impact on each speech.
(d) provide the learners with information about those grammatical structures that have an important impact on each speech and guide the participants to write down the main part of each grammatical structure;	This is based on: (a) the results of Hulstijn and Hulstijn's (1984) study, which suggested that participants with explicit knowledge of the target grammatical structures could apply them better than participants lacking such knowledge; and (b) the results of the second and third pilot studies, which showed that three participants noted only the names of grammatical structures. This practice might not effectively promote the accuracy of the participants' speech because participants may still have to think of the form of each grammatical point as they speak.
and (e) guide the learners to plan the transition words or phrases they will use in their speech.	This is based on the result of the third pilot study, which showed that, in their vocabulary planning, all participants did not plan any transition words or phrases, which were crucial for helping listeners understand their speech.

Table 2. (continued)

Principles	Rationale
<p>3. Instructions that aim to guide L2 learners in planning meaning/form of their speech should:</p> <p>(a) guide the learners to plan the meaning before they plan the form; (i.e., the instructions that guide the learners in planning meaning should precede instructions that guide learners in planning form);</p> <p>and (b) guide the learners to integrate their planning of meaning and their planning of form.</p>	<p>Levelt (1989) proposed that the speech production process proceeds in three stages: first, speakers conceptualize the message. Next, they formulate its language representation. Finally, they articulate the message.</p> <p>Levelt (1989) proposed that when speakers reach the automatic stage, certain steps in the process of their speech production may be processed in parallel. Under the meaning/form-focused pre-task planning condition, participants are provided with instructions that guide them to plan both meaning and form of their speech. These instructions are likely to help them plan both the meaning and the form more effectively; consequently, there is a high possibility that the participants will plan the meaning and the form of their speech in parallel.</p>
<p>4. L2 learners should be guided to consider their listeners' needs while performing their pre-task planning</p>	<p>This assumption is based on Yule, Powers and Macdonald's (1992) study, which showed that paying attention to listeners' needs helps speakers produce more effective speech.</p>
<p>5. L2 learners should be guided to plan their speech in the second language.</p>	<p>Levelt (1989) hypothesized that speakers create a preverbal message that is not linguistic in nature. They then convert this preverbal message into an internal speech that is verbal in nature. This internal and verbal speech is then sent to the articulator, which transforms it into overt speech. In accordance with this hypothesis, the practice of guiding participants to perform their pre-task plans in their second language should be more advantageous than letting them plan in their first language (L1). Since the first practice will encourage participants to convert their preverbal message to internal speech through their second language, the process of transforming their internal speech into overt speech will be facilitated.</p>

Table 2. (continued)

Principles	Rationale
6. L2 learners should be guided to make notes of their plans in short form, using words, phrases, or short sentences and to refrain from writing down everything they want to say in detail.	The results of the first pilot study showed that a participant who wrote down everything he wanted to say in his speech tended to memorize his written plans. Consequently, he recited the plans instead of speaking naturally when making his speech.
7. Specially designed note-sheets should be given to L2 learners to remind and to direct them to follow the suggested pre-task planning instructions.	The results of the second pilot study showed that a specially designed note-sheet, rather than a blank piece of paper, could more effectively lead participants to follow the pre-task planning instructions.

Procedures

Each participant in the MP, FP, and MFP conditions first received a 15 minute-think-aloud training session. They, then, worked one-on-one with the researcher in two separate sessions in which they carried out first the instruction and second the argumentative task. The participants were asked to produce monologues without any other audience present rather than engage in dialogue as it was assumed that they would feel more comfortable and more secure, and would thus be more likely to perform to the best of their ability. They were given 15 minutes to pre-plan for each of these monologues. In addition, they were required to provide plan-aloud reports while performing their strategic planning. They were then given 5 minutes to perform the instruction monologue and 10 for the argumentative monologue. The time allotted to each monologue was greater than they required as, on average, they spent only 2.5 minutes on the instruction monologue and 3.5 minutes on the argumentative monologue. They were given plenty of time to reduce pressure when performing their monologues. After performing each monologue, the researcher engaged the participants in a 10-minute interview. Participants under the MinP condition followed the same procedure, except that they did not receive the 15 minute-think-aloud training session, and were not given time to strategically plan each monologue. The participants' plan-aloud reports, monologues, and retrospective interviews were all audio-recorded.

Data coding and scoring

The data included (1) audio-recorded plan-aloud protocols for the instruction and the argumentative tasks; (2) the strategic plans; (3) audio-recorded speech for the instruction and the argumentative tasks; and (4) audio-recorded retrospective interviews.

The plan-aloud protocol data were first segmented into communicative units (c-units). To establish a coding scheme for the participants' plan-aloud protocols, I relied on the two main principles of the constant comparative method (Glaser & Strauss 1967), "an inductive approach that produces theory grounded in the data" (Lockhart & Ng 1995:614). The first principle was "to allow categories to emerge from the data, rather than imposing preconceived categories on the data" (Lockhart & Ng 1995:614). The second principle was to refine each coding category and identify its properties through comparing all instances coded in one category with other categories (Lockhart & Ng 1995). Following these two principles, I found three main categories of thought processes: (1) communicative goal setting; (2) meaning planning; and (3) form planning. The meaning planning category consisted of five subcategories: (a) generating ideas (GI), (b) organizing discourse (OD), (c) revising ideas (RVI), (d) rehearsing ideas (RHI), and (e) meaning planning procedure (MPP). The form planning category consisted of five subcategories: (a) selecting lexical items (SLI), (b) applying syntactic rules (ASR), (c) revising language (RVL), (d) rehearsing language (RHL), and (e) form planning procedure (FPP) (see definitions of each category and each subcategory with examples from the protocols in Appendix 5).

Participants' application of their strategic plans was assessed through four measures: (1) the number of planned ideas that appeared in actual speech per t-unit (PIPT); (2) the number of unplanned ideas that appeared in actual speech per t-unit (UPIPT); (3) the number of planned grammatical structures that appeared in actual speech per t-unit (PGPT); and (4) the number of unplanned grammatical structures that appeared in actual speech per t-unit (UPGPT).

The task performance data were measured with respect to their accuracy, complexity, and fluency. Speech accuracy was measured by two general measures: (1) the percentage of error-free clauses (PEFC) and (2) the number of errors per 100 words (NER). Speech complexity was assessed through two measures: (1) the number of sentence nodes (s-nodes) per T-unit (SNPT) and (2) the number of clauses per T-unit (CPT). In addition, following Wendel (1997), the imperative forms and auxiliary verbs in each instruction monologue, as well as the comparative and conditional structures in each argumen-

tative monologue, were counted. Speech fluency was analyzed in terms of three measures: (1) speech rate A (SRA) (the total number of syllables in each speech, divided by the total articulation time and multiplied by 60); (2) speech rate B (SRB) (as in SRA but with all repetitions, false starts, replacements, and asides in Thai removed); and (3) the percentage of total pausing time (PTPT).

Data analysis

To analyze the plan-aloud protocols, first, a series of Wilcoxon matched-pairs signed-ranks tests were performed to investigate whether the mean percentages of meaning and form episodes in the plan-aloud protocols under each of the MP, FP, and MFP conditions were significantly different. Then, a series of two-way (repeated-measure) ANOVAs were performed to test the differences among the mean percentages of the episodes for the three main categories and the ten subcategories. These analyses were followed by Tukey's HSD post hoc multiple tests when significance was reached in an ANOVA.

To analyze the data regarding (1) participants' application of their strategic plans and (2) the accuracy, complexity, and fluency of the participants' speech, a series of two-way (repeated-measure) ANOVAs were performed. These analyses were followed by Tukey HSD post hoc multiple tests where appropriate.

Results

Attention under each of the three foci of strategic planning

Table 3 shows that for both the instruction and the argumentative tasks the participants primarily engaged in meaning planning in the MP, FP and MFP conditions. The results of the Wilcoxon matched-pairs signed-ranks test indicated that, for both tasks, the mean percentages of the meaning planning episodes under the three foci of strategic planning were significantly larger than the mean percentages of the form planning episodes ($p = .005$ or $.006$). As a result, Hypothesis 1 was confirmed, whereas Hypotheses 2 and 3 were not supported. Hypotheses 4 and 5 received only limited support. Although the two-way (repeated measure) ANOVAs showed significant effects for both (1) meaning planning episodes ($F = 4.51$, $p = .02$) and (2) form planning episodes ($F = 4.56$, $p = .02$), the Tukey's post hoc tests failed to show significant differences between pairs of the meaning planning and form planning mean scores as predicted in Hypotheses 4 and 5 (see Table 3).

Table 3. Plan-aloud protocol results

Coding categories	Task	Strategic planning conditions			F-value	Sig.	Significant differences	Predicted differences (for both tasks)
		MP	FP	MFP				
		<i>M</i>	<i>M</i>	<i>M</i>				
1. Goal setting	T1	1.29	.74	1.09	0.01	.993	n.s.	none
	T2	2.46	3.00	2.48				
2. Meaning planning	T1	96.38	81.27	88.29	4.51	.020	MP>FP	MP>MFP>FP
	T2	94.26	78.14	79.71			MP>FP	
3. Form planning	T1	2.33	17.99	10.62	4.56	.020	FP>MP	FP>MFP>MP
	T2	3.28	18.86	17.81			FP>MP	
3.1 Selecting lexical items*	T1	.85	10.82	1.43	8.96	.001	FP>MP, MFP	FP>MFP>MP
	T2	1.64	9.87	.23			FP>MP, MFP	
3.2 Revising language*	T1	.27	.00	5.33	6.30	.006	MFP>MP, FP	FP>MFP>MP
	T2	.31	1.38	9.07			MFP>MP	

Key: T1 = Instruction task
 T2 = Argumentative task
 M = Mean percentages of c-units of plan-aloud protocol data
 MP = Meaning-focused strategic planning condition
 FP = Form-focused strategic planning condition
 MFP = Meaning/form-focused strategic planning condition
 n.s. = No significant differences
 = The results of two-way (repeated-measures) ANOVAs on mean percentages of “selecting lexical items,” and “revising language” episodes are included in Table 3 because the ANOVAs indicated significant strategic planning effects.

Application of strategic plans

Hypotheses 6 and 7 also received only limited support. The two-way (repeated measure) ANOVAs showed significant foci of strategic planning effects for the PIPT ($F = 6.66, p = .004$); UPIPT ($F = 11.98, p = .000$), and PGPT ($F = 10.44, p = .000$) measures, but not for the UPGPT measure. In addition, the Tukey’s post hoc tests failed to show significant differences among some pairs of the PIPT, UPIPT, PGPT mean scores as predicted in Hypotheses 6 and 7 (see Table 4).

Effects of foci of strategic planning on quality of speech

Hypothesis 8 received only limited support, while Hypotheses 9, 10, and 11 were not supported (see Table 5). Even though the two-way (repeated measure) ANOVAs showed significant planning effects for (1) all the accuracy measures (PEFC: $F = 5.85, p = .002$; NER: $F = 5.54, p = .003$), (2) all the complex-

Table 4. Results for the application of strategic plans

Measure	Task	Strategic planning conditions			F-value	Sig.	Significant differences	Predicted differences (for both tasks)
		MP	FP	MFP				
		<i>M</i>	<i>M</i>	<i>M</i>				
PIPT	T1	.67	.55	.91	6.66	.004	MFP>MP, FP MP>FP	MFP>MP>FP
	T2	.86	.50	.81				
UPIPT	T1	.33	.45	.19	11.98	.000	MFP<FP MFP, MP<FP	MFP<MP<FP
	T2	.24	.50	.19				
PGPT	T1	.26	.35	.40	10.44	.000	MFP>MP, FP MFP>MP	MFP>FP>MP
	T2	.22	.32	.56				
UPGPT	T1	.14	.19	.15	1.19	.321	n.s.	MFP<FP<MP
	T2	.15	.25					

Key: T1 = Instruction task
 T2 = Argumentative task
 M = Mean percentages
 MP = Meaning-focused strategic planning condition
 FP = Form-focused strategic planning condition
 MFP = Meaning/form-focused strategic planning condition
 PIPT = Percentage of planned ideas used in produced speech per t-unit
 UPIPT = Percentage of unplanned ideas used in produced speech per t-unit
 PGPT = Percentage of planned grammatical structures used in produced speech per t-unit
 UPGPT = Percentage of unplanned grammatical structures used in produced speech per t-unit
 n.s. = No significant differences
 = Mean percentages of PGPT and UPGPT do not add up to 100 because only the target grammatical structures, not all grammatical structures, were counted.

ity measures (SNPT: $F = 5.04$, $p = .005$; CPT: $F = 4.13$, $p = .013$) and (3) all the fluency measures (SRA: $F = 3.76$, $p = .019$; SRB: $F = 3.70$, $p = .02$; PTPT: $F = 5.31$, $p = .004$), the Tukey's post hoc test indicated that the MFP, MP, and FP conditions, for both tasks, were not significantly different in promoting the accuracy, complexity, and fluency of speech. In addition, the post hoc tests revealed selective effects for the MFP, MP, and FP conditions over the MinP condition in promoting (1) speech accuracy (MFP for the instruction task; MFP, MP, and FP for argumentative task), (2) speech complexity (MP for the instruction task; MFP for the argumentative task), and (3) speech fluency (MFP, MP, and FP for the instruction task; FP for the argumentative task).

Table 5. Results for task performance

Measures	Task	Strategic planning conditions				F-value	Sig.	Significant differences (for both tasks)	Predicted differences
		MP	FP	MFP	MinP				
		<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>				
Complexity									
SNPT	T1	0.60	0.96	0.71	0.77	5.04	.005	MP>MinP	MFP>MP
	T2	1.64	2.02	2.39	2.73			MFP>MinP	>FP>MinP
CPT	T1	1.34	1.60	1.47	1.53	4.13	.013	n.s.	MFP>MP
	T2	2.09	2.37	2.49	2.76			MFP>MinP	>FP>MinP
Accuracy									
PEFC	T1	48.00	69.23	69.24	75.39	5.85	.002	MFP>MinP	MFP>FP>
	T2	45.14	61.47	61.36	65.89			MFP, MP, FP>MinP	MP>MinP
NER	T1	7.67	4.58	4.85	3.25	5.54	.003	MFP<MinP	MFP<FP
	T2	8.95	6.59	7.12	6.16			MFP<MinP	<MP<MinP
Fluency									
SRA	T1	65.26	89.10	95.34	81.73	3.76	.019	FP>MinP	MFP>FP
	T2	51.29	71.48	81.93	70.83			FP>MinP	>MP>MinP
SRB	T1	57.79	79.47	85.74	67.14	3.70	.020	FP>MinP	MFP>FP
	T2	42.68	59.36	70.60	49.15			FP>MinP	>MP>MinP
PTPT	T1	30.69	13.72	9.08	18.88	5.13	.004	MFP, MP, FP<MinP	MFP<FP
	T2	37.27	27.92	20.86	33.80			n.s.	<MP<MinP

- Key: T1 = Instruction task
 T2 = Argumentative task
M = Mean scores
 MP = Meaning-focused strategic planning condition
 FP = Form-focused strategic planning condition
 MFP = Meaning/form-focused strategic planning condition
 MinP = Minimal strategic planning condition
 SNPT = Sentence nodes per t-unit
 CPT = Clauses per t-unit
 PEFC = Percentage of error free clauses
 NER = Number of errors per 100 words
 SRA = Speech rate A
 SAB = Speech rate B
 PTPT = Percentage of total pausing time

Discussion

Overall, the learners chose to focus on meaning when they engaged in strategic planning, irrespective of the intended planning foci. However, differences

among the groups were evident. The quantitative results together with a qualitative analysis of the planning protocols indicate that there was (1) intensive meaning planning, accompanied by superficial form planning in the case of the MP, (2) extensive vocabulary planning in the case of the FP, and (3) numerous grammatical corrections during the formulating stage in the case of the MFP. Overall, the instructions for the MP, FP, and MFP conditions had the hypothesized effect of respectively guiding participants (1) to direct their attention toward the planning of meaning, (2) to direct their attention toward the planning of (lexical) form; and (3) to balance their attention between the planning of meaning and the planning of form.

The present study suggests that the PIPT, UPIPT, PGPT, and UPGPT, which this author developed for this study, are effective for explaining L2 learners' application of strategic plans. In addition, the present study reveals (1) positive effects of the MFP condition (for the instruction task), and the MP condition (for the argumentative task) in promoting application of planned meaning and (2) positive effects of the MFP condition in promoting application of planned form (for both tasks). These findings can be explained as follows.

The positive effects of the MFP condition in promoting application of planned meaning in the instruction task may reflect the fact that the MFP participants' plans contained sufficient planned ideas for conveying the essential information and a clear organisation. The MFP condition may have guided the participants to achieve a balanced orientation between elaborated meaning and accurate form since it provided them with instructions on how to plan both meaning and form for their speech. Having this balanced goal may have led the MFP participants to plan only the essential ideas. The meaning plans of the MFP participants were found to be as clear as those of the MP participants and clearer than those of the FP participants. These two factors increased the likelihood of the MFP participants being able to use their planned ideas when they performed the instruction task. However, in the argumentative task, the MFP participants, when compared to the MP participants, allocated more attention to the planning of form and consequently paid less attention to developing ideas. This is borne out by the think-aloud protocols, which showed that the MFP participants allocated 17.81 percent of their attention, nearly as much as the FP participants (18.86 percent), toward planning form, while the MP participants allocated only 3.28 percent of their attention towards form (see Table 3).

The positive effects of the MP condition in promoting the application of planned meaning for the argumentative task may reflect the participants'

concern for more elaborated meaning. In the argumentative task, the MP condition may have led the participants to focus on propositional meaning rather than on form and, as a result, they used whatever grammatical structures came to mind when they performed the task. In the case of the instruction task, however, the MP participants may have planned too many ideas and thus were forced to discard some when performing the task. For both tasks, the positive effects of the MFP condition in promoting the application of planned form may reflect the participants' attempts to use the grammatical structures that were suggested in the instructions in their plans, and, later, to apply these grammatical structures in their speech. Moreover, the MFP participants, when compared to the MP and FP participants, may have had clearer ideas about the relationship between form and meaning since the MFP condition had instructed the participants to integrally plan for both. These two factors may have enabled the MFP participants to use a variety of planned grammatical structures in their speech.

Thus, the effects of foci of strategic planning in promoting participants' application of their strategic plans may depend on the ability of such foci to influence the participants to (1) balance attention between meaning and form, (2) produce clear and economical meaning plans and (3) perceive the relationship between meaning and form. Strategic planning directed at combined meaning/form appears to be relatively more effective than planning that is focused separately on meaning or form.

The findings of the the present study show (1) positive effects for the MFP condition (for the instruction task) and (2) for the MFP, MP and FP conditions (for the argumentative task) on accuracy. These findings may be explained as follows. The MFP condition appears to have oriented the participants to the need for accuracy in both tasks. Thus, the MFP participants, when compared to the MinP participants, set aside more attentional resources for monitoring grammatical accuracy while planning and in their subsequent speech. Also, as is evident in the the plan-aloud protocols, the MFP participants paid more attention to grammatical accuracy in their strategic plans than did the participants in the MP condition for both tasks and, more suprisingly, than the FP condition for the argumentative task (see the subcategory 'revising language' in Table 3). The second finding can be explained as follows. The MFP, MP and FP conditions, when compared to the MinP condition, may have significantly decreased the processing load on the conceptualizer, the formulator, or on both. Since the processing load on the conceptualizer and/or the formulator were reduced, the MFP, MP and FP participants, when compared to the MinP participants, may have made more attentional resources available for monitoring

the grammatical accuracy of their speech in the case of the more demanding argumentative task. In contrast, no differences emerged for the easier instruction task because this posed fewer demands on their processing capacity.

The findings of the present study regarding speech accuracy are in line with those of Foster and Skehan (1996, 1997), Mehnert (1998), and Wigglesworth (1997). Some other studies, however, have reported a non-significant effect of strategic planning on accuracy (Crookes 1989; Ortega 1999; Ting 1996; Wendel 1997; Williams 1992). Wendel (1997) argued that the non-significant effects of strategic planning in promoting speech accuracy were due to the fact that “‘off-line’ monitoring cannot affect ‘on-line’ monitoring” (p. 142). However, the findings of the present study suggest that strategic planning can have a positive effect on accuracy. Strategic planning may have (1) influenced the participants to orient themselves towards accurate form and (2) reduced the processing loads on the conceptualizer and/or formulator so that the learners had sufficient attentional resources to monitor grammatical structures during task performance. It is possible that the non-significant effects of strategic planning in promoting accuracy reported in some studies may be attributed to the failure of the strategic planning conditions used to create the two crucial conditions needed to promote accuracy: 1) the participants’ orientation toward accurate form and 2) the availability of sufficient attentional resources for grammatical monitoring. This may have arisen because the researchers failed to provide their participants with effective instructions for planning strategically.

The present study found positive effects for the MP condition (for the instruction task) and for the MFP condition (for the argumentative task) on complexity. These findings can be explained as follows. First, the MP and MFP conditions may have reduced the processing load on the conceptualizer. This seems evident from the results of the analysis of the PIPT and PGPT measures, which showed that the MP participants used a large number of planned ideas in their speech, while the MFP participants used a large number of both planned ideas and planned grammatical structures in their speech (see Table 4). Second, the two conditions may have influenced the participants to orient themselves toward more elaborated meaning. This is supported by the analysis of the plan-aloud protocols, which demonstrated that the MP and MFP participants allocated, respectively, more than 94 percent and 79 percent respectively of their attention to planning the meaning of their speech (see Table 3). Finally, the MFP condition may have enabled the participants to use complex grammatical structures in their strategic plans for the argumentative task. This is possible because the condition provided the participants with informa-

tion about the form and use of grammatical structures that were relevant to the assigned task.

The present study reveals the positive effects of the MFP, MP, and FP conditions (for the instruction task) and the FP condition (for the argumentative task) in promoting speech fluency. Again, these findings may reflect the fact that the MFP, MP, and FP conditions, when compared to the MinP condition, reduce the processing load on the conceptualizer and/or formulator. The lighter processing loads may have enabled the conceptualizer and/or formulator to operate in parallel and, as a result, the participants produced more fluent speech. The findings for fluency are consistent with those of most previous studies (Foster & Skehan 1996, 1997; Mehnert 1998; Ortega 1995, 1999; Wendel 1997; Wigglesworth 1997). These studies have indicated that strategic planning promotes speech fluency because it facilitates conceptualisation. The present study also suggests that planning form as well as meaning can promote fluency.

It is important to note that, although the findings discussed in the preceding sections suggest that each of the three foci of strategic planning is more effective than the MinP condition in promoting complexity, accuracy, and fluency of speech, L2 learners do best when they engage in strategic planning that is focused jointly on meaning and form. This study has shown that the MFP condition, when compared to the MP and the FP, had the greatest effect on the quality of the speech produced.

Conclusion

The present study supports the findings of previous research regarding strategic planning. The most important contribution of this study is that it provides L2 learners and L2 educators with a clear explanation of how the three different foci of strategic planning effected the L2 learners' (a) cognitive strategic planning processes, (b) their application of strategic plans and (c) the quality of their speech. Moreover, it has offered an explanation for the mixed results regarding speech accuracy in previous research.

The present study has implications both for pedagogy and for research. In terms of pedagogical practice, the findings of this study suggest that L2 learners should attempt a meaning/form focus in strategic planning. This is because it is more effective than either the strategic planning of meaning or the strategic planning of form in promoting (1) an optimal balance of attention between the planning of meaning and the planning of form; (2) the implementation of

strategic plans; and (3) a balanced quality of speech. In addition, the findings of the present study suggest that this author's principles for developing instructions for strategic planning and the note-sheets given to the participants (see Table 2) are an effective means of guiding strategic planning.

Nevertheless, three limitations of this study should be considered. First, the plan-aloud protocols may have failed to reveal some of the cognitive processing steps involved in strategic planning, particularly if the steps occurred too quickly for the participants to articulate all the processes they engaged in. Second, the study recruited students with intermediate English proficiency. The positive effects of the foci of strategic planning that the present study has demonstrated may not be generalizable to L2 learners with different proficiency levels. Finally, the participants in the present study were required to produce monologues. Consequently, the research findings may not be generalizable to situations in which participants are required to engage in dialogic discourse.

Appendix 1

Meaning/form planning guidelines for the instruction task

MF.1

Guidelines for the Instruction Task

Follow these steps to plan your Instruction speech:

1. Consider* your friend's knowledge of your hometown.
2. Visualize* the two procedures* that you will explain to your friend:
 - a) how your friend will get from the train station to your school;
 - b) when and where she/he will meet you.
3. Break each procedure into short steps.
4. Think how you will arrange your instructions so that it is easy for your friend to follow them.
5. Order each step in the correct sequence*.
6. Also, do the followings:
 - a) Think of all words you want to use in your message, and note only one word for one meaning.
 - b) Think of transition words or phrases, such as *first, second, next . . . finally* that will connect your instructions so that it is easy for your friend to follow them.
 - c) Think of grammatical structures that play an important role in the task and write down the main parts of the grammatical structures.

(continue)

-
- Note.**
1. consider = □□□□□□
 2. procedure = □□□□□□□□□□□□
 3. sequence = □□□□□□□□□□
 4. visualize = □□□□□□

Appendix 2

Meaning/form guidelines for the argumentative task

MF.2

Guidelines for the Argumentative Task

Follow these steps to plan your argumentative speech:

1. Decide what your position* is on the topic, and state it in your introduction, so that your listeners know what you are going to say about the topic.
2. Choose two or three main reasons to support your position.
3. Support each main reason with details or examples.
4. Think how you will arrange your information, so that it is easy for your listeners to understand your argument.
5. Also, do the followings:
 - a) Think of all words you want to use in your message, and note only one word for one meaning.
 - b) Think of transition words or phrases, such as *first, second, next ... finally* that will connect your information so that it is easy for your listener to follow your speech.
 - c) Think of grammatical structures that play an important role in the task and write down the main parts of the grammatical structures.

(continue)

-
- Note.** 1. position = □□□□□□
 2. state = □□□□□□□□□□□□

Grammatical structures that are common and needed for the argumentative speech are:

- comparative structures
- conditional forms
- Present Simple tense

For example, you might note down the main parts of grammatical features as follows:

Grammatical structures
Your notes

-er... than
more... than
less... than
not as... as

as... as

easier than,
more difficult than,
less difficult than,
as difficult as
not as difficult as

conditional forms

If I exercise, I will lose weight.
If I exercised, I would lose weight.
If I were a bird, I would fly.

6. Try to join content, vocabulary and grammar planning. For example, you might first think of a sentence you want to say. Then you choose the words and grammatical structures you want to use to make the sentence.
7. Plan your conclusion by restating your position.
8. Please note down your plan in English, but do not write out everything in detail. You then have to talk without your notes.

Appendix 3
Meaning and form planning note-sheet for the instruction task

	MF.1
<u>Note-sheet for the instruction task</u>	
Use this note-sheet to prepare your speech.	
1) Consider your friend's knowledge of your hometown	
2) Visualize:	
I) how your friend will get from the train station to your school; and	
II) when/where she/he will meet you.	
3) Break each procedure into short steps.	
4) Think how you will arrange your instructions so that it is easy for your friend to follow them.	
5) Order the steps in the correct sequence.	
6) Try to join content, vocabulary, and grammar planning.	
I. _____	
A. _____	
B. _____	
C. _____	
D. _____	
E. _____	
F. _____	
II. _____	
A. _____	
B. _____	
C. _____	
D. _____	
E. _____	
F. _____	
7) Plan your opening.	

8) Plan your closing.	

(Adapted from Stein, W. (1991). *Communicative skills that work: A functional approach for life and work, Book 1*. Chicago: Contemporary Books, Inc. pp. 26-28).

Appendix 4

Meaning and form planning note-sheet for the argumentative task

	MF.2
<u>Note-sheet for the argumentative task</u>	
Use this note-sheet to prepare your speech.	
1) State your position.	<hr/> <hr/>
2) Try to join content, vocabulary, and grammar planning.	<hr/> <hr/>
3) Choose two or three main reasons to support your position and think how to arrange your information, so that it is easy for your listeners to understand your argument.	
I.	<hr/>
II.	<hr/>
III.	<hr/>
4) Support each main reason with details or examples.	
I.	<hr/> <hr/>
A.	<hr/> <hr/>
B.	<hr/> <hr/>
II.	<hr/> <hr/>
A.	<hr/> <hr/>
B.	<hr/> <hr/>
III.	<hr/> <hr/>
A.	<hr/> <hr/>
B.	<hr/> <hr/>
5) Plan your conclusion.	<hr/> <hr/> <hr/>

(Adapted from Kayfetz, J. L. & Stice, R. L. (1987). *Academically Speaking*. Massachusetts: Heinle & Heinle Publisher, p. 91)

Appendix 5

Categories and sub-categories of the plan-aloud protocol coding scheme

Coding categories	Definitions and examples from the protocols
1. Communicative goal setting	<p>Protocol statements under this category show that participants are setting communicative goals for their speech. This is evident in the form of reading and elaborating on task instructions, reviewing, reinstating, reconsidering, or evaluating their communicative goals.</p> <p>Examples: “My friend will come. I cannot pick her up at the train station because I have a test. I will leave her a message.” (P27/LFPP/T1/Cu1)</p>
2. Meaning planning	
2.1 Generating ideas	<p>Protocol statements under this category indicate that participants are attending to the content of the speech. This is evidence in the form of considering and/or formulating main or supportive ideas for the speech.</p> <p>Examples: “First, I will call my friend to tell her that I can’t go, telephone. Hello hello I can’t meet you at the railway station.” (P12/CFPP/T1/Cu1)</p>
2.2 Organizing discourse	<p>Protocol statements under this category indicate that participants are considering or organizing the content of their speech at the levels beyond the sentence. This is evident in the form of individually or collectively considering the coherence of speech content in reference to goal and discourse organization, or considering cohesive devices that make links between two or more sentences/content.</p> <p>Example: “ First, I should introduce myself.” (P14/CFPP/T1/Cu3)</p>
2.3 Revising ideas	<p>Protocol statements under this category indicate that participants are changing their planned ideas. This is evident in the form of reviewing and reformulating main or supportive ideas in their plans.</p> <p>Example: “ When you arrive my school, we should meet at Rachawadee, Rachawadee Hall, in front of Rachawadee. Oh, I will make a change. At the information room, information room under build B-U-I-L-D, the seventh building, at the seventh building.” (P21/LFPP/T2/Cu10-Cu12)</p>

Coding categories	Definitions and examples from the protocols
2.4 Rehearsing ideas	<p>Protocol statements under this category indicate that participants are reciting the ideas in their plans or practising their speech following the ideas in their plan.</p> <p>Example: “I should check again. Hello, I’m sorry that I can’t receive you at the railway station.” (P15/CFPP/T1/Cu17)</p>
2.5 Meaning planning procedure	<p>Protocol statement under this category indicates that participants are focusing on the procedure of their “meaning planning.”</p> <p>Examples: “I could not tell all the details.” (P14/CFPP/T1/Cu8) “I have planned when to meet, and where to meet. I should end the plan.” (P12/CFPP/T1/Cu16-Cu17)</p>
<p>3. Form planning</p>	
3.1 Selecting lexical items	<p>Protocols statements under this category show that participants are selecting lexical units for their messages. This is evident in the form of considering the appropriateness, quality or desirability of lexical units or transition words or phrases, and selects the right one for a particular conceptual message; evaluation of their lexical selection; or solving lexical problems.</p> <p>Examples: “..... you see you will see many service cars, tuk-tuk. I should use ‘taxi’ instead.” (P39/CLFPP/T1/Cu5-Cu6) “What is an English word for ‘discipline’? I should find another way to say this word, does not respect, does not respect rules, does not respect a rule.” (P33/CLFPP/T2/Cu6)</p>
3.2 Applying syntactic rules	<p>Protocol statements under this category indicate that participants are applying syntactic rules, evaluating the accuracy of their application of syntactic rules, or focusing on sentence formation.</p> <p>Examples: “... Oh, I should work with grammatical structures. It is easier. You don’t have to worry, easier than.” (P22/LFPP/T2/Cu13-Cu14)</p>
3.3 Revising language	<p>Protocol statements under this category indicate that participants are changing their “planned form.” This is evident in the form of reviewing and changing lexical units, transition words or phrases, or grammatical structures in their plans.</p>

Coding categories	Definitions and examples from the protocols
3.3 (continue)	<p>Example:</p> <p>“Change ‘ugly’, ‘ugly’ to ‘not beautiful’”. (P18/CFPP/T2/Cu21)</p> <p>“If I ... if I wear coat in sunny Thailand, if I wear coat I should use ‘wore’. If I wore coat in sunny Thailand, I what? I would be ... I would be mad.”</p> <p style="text-align: right;">(P25/LFPP/T2/Cu9)</p>
3.4 Rehearsing language	<p>Protocol statements under this category indicate that participants are reciting or practicing the vocabulary units or grammatical structures in their plans</p> <p>Examples:</p> <p>“The vocabulary are: Yorwor, in front of, motorcycle, school, me, along, then, of, when.”</p> <p style="text-align: right;">(P4/LP/T1/Cu31)</p> <p>“Use transition words, use if-clause, use and first finally and then for and.”</p> <p style="text-align: right;">(P6/LP/T2/Cu19)</p>
3.5 Form planning procedure	<p>Protocol statements under this category indicate that participants are attending to their processes of selecting lexical units, transition words or phrases, or applying syntactic rules. This is evidence in the form of guiding themselves on how to select lexical units, or transition words or phrases, how to apply syntactic rules, and how to revise lexical units, transition words, phrases, or syntactic structures in their plans.</p> <p>Examples:</p> <p>“There must be transition words, or phrases.”</p> <p style="text-align: right;">(P2/LP/T1/Cu6)</p> <p>“I should better not use this word”.</p> <p style="text-align: right;">(P6/LP/T2/Cu29)</p> <p>“... Oh, I should work with grammatical structures.”</p> <p style="text-align: right;">(P2/LP/T2/Cu13)</p>

Key: The English translation of protocols that were spoken in Thai appears in normal type. Protocols that were spoken in English appeared in bold type.

P = Participant

T1 = Instruction task

T2 = Argumentative task

Cu = C-unit

MP = Meaning-focused strategic planning condition

FP = Form-focused strategic planning condition

MFP = Meaning/form-focused strategic planning condition

... = Three dots indicate a pause that is shorter than 3 seconds

..... = Six dots indicate a pause that is 3 seconds, or longer than 3 seconds

CHAPTER 5

The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency

Chieko Kawauchi

Kurume University, Japan

Introduction

The purpose of the study reported in this chapter is to investigate the effects that strategic planning has on L2 learners' performance of an oral narrative task. The focus is on the role that proficiency plays in the effect of planning and also on the type of planning, which is operationalised in terms of three pre-task activities – Rehearsal, Writing, and Reading.

Many studies have found that strategic planning significantly facilitates fluency in L2 oral production (Crookes 1989; Foster & Skehan 1996; Mehnert 1998; Ortega 1995, 1999; Wigglesworth 1997). Language complexity also increases, especially 1) for more proficient learners (Wigglesworth 1997), and 2) with more cognitively demanding tasks (Foster & Skehan 1996). When it comes to accuracy, however, the effects of strategic planning are less certain. Wigglesworth (1997), for example, showed that planning had only a limited effect on verb morphology accuracy and only with high proficiency learners. Thus, the effects of planning on accuracy are not clear-cut and appear to be influenced by specific task types, the choice of measures for analysis, and the learners' proficiency levels.

Previous studies suggest that the role of planning is three-fold. First, it eases the on-line processing load as well as reducing communicative stress to yield higher fluency. Second, planning helps learners to access their maximal level of lexical and structural knowledge, which, in turn, will enable them to use more complex language. Third, it facilitates the allocation of conscious attention to form and thus helps learners to generate more accurate language.

However, whereas several studies have considered how the effect of planning varies according to the cognitive load imposed by the task there has been almost no consideration of the interaction between proficiency and strategic planning in the effect the latter has on task performance. Indeed, these studies have investigated learners with a very limited range of proficiency. They have examined mainly intermediate learners and post-beginners (Ellis 1987), intermediate learners (Ortega 1995), pre-intermediate learners (Foster & Skehan 1996; Skehan & Foster 1997), and early intermediate learners (Mehnert 1998).

A few studies have investigated learners with mixed levels of proficiency. For example, Crookes (1989) carried out an experiment on a wide range of Japanese learners of English, whose TOEFL scores ranged from 460 to 620. However, it is difficult to see how the higher and lower proficiency learners responded to the tasks, since no distinction was made in his analyses. To the best of this author's knowledge, the study by Wigglesworth (1997) is the only study that took proficiency into account. Focusing on 28 high and 23 low proficiency learners, Wigglesworth examined the effect of one-minute of planning time on a tape-mediated oral test consisting of four kinds of tasks (i.e., telephone answering machine message; picture description; summary of conversation; and general discussion). The findings indicated that the planning time only helped the more highly proficient learners to produce more complex language (i.e., subordinate clauses) and more accurate language (i.e., verb morphology). This was particularly true in the case of the tasks with a high cognitive load such as the picture description task. The opportunity to plan did not seem to benefit learners at lower levels of proficiency. As for fluency, lower proficiency learners produced fewer self-repairs than high proficiency learners in all four tasks, while higher proficiency learners improved only in the summary task, which was the most difficult. However, this study does not permit definite conclusions regarding the effect of planning time on fluency, since one group apparently monitored more than the other group. The findings from this study led Wigglesworth to conclude that "for the high-proficiency candidates, planning time may improve accuracy on some measures where the cognitive load of the task is high, but that this effect does not extend to the low-proficiency candidates" (p. 85).

This finding is important in that it suggests that the effects of planning will differ according to the learner's proficiency level. Specifically, the effects on complexity and accuracy are more likely to be found in higher proficiency learners when the task is cognitively more demanding. A similar suggestion is made by Ortega (1999), who examined the retrospective reports of learners (see also Chapter 3). The participants were American learners of Spanish

who were considered to be “at an advanced level of oral language ability” (p. 121). The learners’ reports indicated several advantages for planning – it decreased communicative stress and the perceived difficulty of the task while making notes helped the learners to remember what to say. In addition, the reports also indicated a number of moderating factors, such as task simplicity and a proficiency ceiling effect. For example, some participants reported that the stories in the task were too simple to make planning necessary. Other participants reported lexical compensation strategies and difficulties with lexical retrieval (e.g., “what you don’t know, you can’t remember, no matter how much extra time you’re given”). Based on these retrospective reports, Ortega claimed that “planning may be more likely to have an effect on the quality of the linguistic output with higher levels of proficiency” (p. 137). Ortega also suggested that the trade-off hypothesis (Foster & Skehan 1996) and the lack of an effect for planning on accuracy reported in previous studies should be re-considered to take account of learners’ proficiency. The implication behind this seems to be that learners with higher level of proficiency might produce more accurate language when given the chance to plan strategically and, in such a case, no trade-off effect between accuracy and complexity will be evident.

These claims from Wigglesworth (1997) and Ortega (1999) are noteworthy. However, further research is necessary to confirm them, since Wigglesworth’s study is based on just one-minute of planning in a testing situation, which is different from most other studies, which were conducted in classrooms or laboratory settings and which allowed ample time for planning, usually 10 minutes. Ortega’s study did not specify how high learners’ proficiency needed to be for a planning effect to become evident

The present study focuses on three different proficiency levels to show more clearly the role of proficiency in any effect for planning. It also attempts to specify the content of planning as a pre-task activity. Most of the previous studies have attempted to examine planning effects by providing a certain amount of time, 10 minutes in most of the cases, and by requiring learners to make notes to ensure that they were engaged in planning. However, it was not clear what the learners were actually doing during this period because what learners did during the planning time was often left to the individual learners themselves. Only a few studies have specified the content of planning in the pre-task activity: Ellis (1987) employed writing prior to the task, Bygate (1996) used repetition of the task, and Sangarun (see Chapter 4) examined the differential effects of planning directed at meaning, form and meaning/form combined.

In this study three kinds of planning activities were chosen: writing a draft (Writing), rehearsing (Rehearsal), and reading a model L2 input (Reading).

These activities frequently arise when learners are given time to plan and are left to their own devices. Ortega (1999; Chapter 3) showed that a majority of learners reported that they employed a range of strategies such as “running through it a couple of times”, “talking to oneself”, or “reading to oneself” (p. 127). These are examples of Rehearsal, although not necessarily oral rehearsal, unless learners are required to speak out loud. Many of the previous studies used note-making. This is close to Writing. Reading related material (L2 input) was also included in the present study to investigate the relationship between planning and noticing/filling the gap. When people plan something, they try to rehearse what they are going to say, often writing it out, and if there is a gap in their knowledge, they will look to fill this by reading related material (cf. Prabhu’s (1987) idea of ‘borrowing’).

These strategic planning activities have a theoretical basis. Writing and Rehearsal, which involve output, can find theoretical support in Swain’s “Output Hypothesis”. When engaged in these activities, “learners need to be pushed to make use of their resources; they need to reflect on their output and consider ways of modifying it to enhance comprehensibility, appropriateness and accuracy” (Swain 1993: 160–161). Also, learners may attempt to attain higher fluency and accuracy by means of a reduction strategy when they have difficulty in accessing the target form (Faerch & Kasper 1983). The Reading activity, which is based on L2 input, relates to “noticing” (Schmidt & Frota 1986; Schmidt 1990) and “focus on form” (Long & Robinson 1998). It is assumed that learners who engage in the Reading activity will notice the gap between their IL and the target language and will interact with the relevant meaning and linguistic information in the input, leading them to focus on form.

There is, however, an important difference between Writing and Rehearsal on the one hand and Reading on the other. It is likely that when learners are engaged in the Reading activity, they can draw on the linguistic forms in the input, in addition to their own IL. On the other hand, when engaged in Writing and Rehearsal, learners are dependent on their own IL resources, since no other information is available to them. Therefore, these three types of activities are hypothesized to produce different effects on the subsequent performance, which will be reflected in measures of fluency, complexity, and accuracy.

Taking these two aspects (i.e., proficiency and planning type) into account, the present study addresses the following two research questions:

- (1) What effect does proficiency have on L2 learners’ performance of an oral narrative task after strategic planning?

- (2) To what extent is the effect of proficiency on L2 learners' performance of an oral narrative dependent upon the type of strategic planning?

Method

Participants

Japanese learners of English with different proficiency levels participated in the study: there were 16 low-intermediate EFL learners (Low EFL), 12 high-intermediate EFL learners (High EFL), and 11 advanced ESL learners (Advanced ESL). The Low EFL learners were all majoring in International Politics in Kurume University, Japan, and they ranged in age from 19 to 21 years. The average length of their English study at the time of the experiment was 7.5 years. Four learners had passed the second level, and the rest of them had passed the pre-second level of the STEP test (the Society for Testing English Proficiency authorised by the Japanese Ministry of Education). These levels are approximately equivalent to 44–50 in the Secondary Level of English Proficiency (SLEP) Test or 420 – 480 in TOEFL (Ogawa 1990).

Both the High EFL and Advanced ESL learners were enrolled in Lancaster University, U.K., at the time of the study. The High EFL learners were students from the Junior Year Abroad Program and registered in various departments after studying an intensive pre-session program at the Institute for English Language Education. Their English proficiency averaged TOEFL 545, ranging from 510 to 580, and the average IELTS was 6 with a range of 5.5 to 6.5. Their average age was 22.0, ranging from 21 to 25. Most of their English study had been completed in Japan, and none had ever studied in English speaking countries before. The Advanced ESL learners were enrolled in Lancaster University as full time students either in undergraduate or graduate programs. Their average English proficiency was TOEFL 588, ranging from 550 to 610, and the average IELTS 6.7 with a range of 5.5 to 7.0. They were in the age range of 21 to 29, averaging 25.2. All of them had been staying in Britain or the USA for more than one year at the time of the investigation.

Materials

The study made use of a narrative task rather than an interactive task. Foster & Skehan (1996, 1999) found that in their studies, which included interactive tasks, there were a few students who were too inactive to be included in

data analysis. Learners' outcomes are also influenced by the interlocutor's re-confirmations, clarification requests, and recasts, which will make it difficult to analyse the effects of planning on individual learners' task performance. In contrast, a narrative task is considered to be cognitively more demanding than personal story telling (Skehan & Foster 1995, 1997; Foster & Skehan 1996; Kawauchi 1998; Wigglesworth 1997), and therefore, is more likely to reveal the effects of strategic planning. Story-elicitation procedures can also prevent too much individual variation in the story lines (Ortega 1999). Ejenberg (1992) showed that her participants reported that a narrative task was more efficient in assessing their oral ability than a dialogue task.

In this study, three different sets of pictures (i.e., Library, Jogging, and Hiking) were used in order to minimize the practice effect resulting from repetitive use of a single picture set. All the picture sets were chosen from the pre-first level of the STEP test and are considered to be similar in both difficulty level and story content (i.e., they depict people who are annoying and who are annoyed). Each picture set consists of a series of four pictures.

Design

The study employed a "within subjects" design, in which learners completed both the unplanned and planned tasks, rather than a "between subjects" design, in which learners were assigned to either an unplanned or planned task. By using a "within subject" design, language production under both planning and non-planning conditions can be compared to reveal differences between conditions more clearly. The basic design of the study is shown in Figure 1.

First, learners performed a task in the unplanned condition (Unplanned Task). In this stage, learners were required to make the best use of their resources in on-line planning. It was expected that the learners would focus mainly on meaning rather than on form (VanPatten 1990) but that they would notice the gap between what they did not know or knew only partially and what was needed. The Unplanned Task served to provide baseline data to compare with the data collected in the subsequent Planned Task.



Figure 1. Design of the study

Table 1. Data collection for the study

	Group 1	Group 2	Group 3
Week 1	Library task (Cycle 1) 1. Unplanned task 2. Planning: Reading 3. Planned task	Hiking task (Cycle 1) 1. Unplanned task 2. Planning: Rehearsal 3. Planned task	Jogging task (Cycle 1) 1. Unplanned task 2. Planning: Writing 3. Planned task
Week 2	Jogging task (Cycle 2) 1. Unplanned task 2. Planning: Rehearsal 3. Planned task	Library task (Cycle 2) 1. Unplanned task 2. Planning: Writing 3. Planned task	Hiking task (Cycle 2) 1. Unplanned task 2. Planning: Reading 3. Planned task
Week 3	Hiking task (Cycle 3) 1. Unplanned task 2. Planning: Writing 3. Planned task	Jogging task (Cycle 3) 1. Unplanned task 2. Planning: Reading 3. Planned task	Library task (Cycle 3) 1. Unplanned task 2. Planning: Rehearsal 3. Planned task

In the second stage, learners were separately assigned to one of the three Planning Activities. Since they already know the content of the picture story, it is highly likely that their main focus in this stage would be on attending to form and also on resolving problems they had in the Unplanned Task. In the third stage, after the Planning Activity, the learners completed a questionnaire designed to provide information about their planning. The questionnaire was expected to minimise the immediate practice effect of repeating the task in the next stage.

Finally, the learners performed the same task again (Planned Task). Since the research questions focus on how each of the three Planning Activities (Writing, Rehearsal, and Reading) affects subsequent task performance, the learners performed each planning activity with a separate task in three weekly sessions. Thus, they performed a total of three tasks, each task consisting of one cycle (i.e., Unplanned – one of the pre-task activities – questionnaire – Planned), as shown in Figure 1. In this way, the present study combines Bygate's (1996; Chapter 2) task repetition with a planning activity.

Procedures

The data for the Low EFL learners were audio-recorded in a language laboratory as part of a regular English class, and those for the High EFL and Advanced ESL learners were collected individually in the author's office. The study was conducted in three weekly sessions as shown in Table 1.

First, learners were asked to describe the set of pictures assigned to them without any preparation. Two minutes were provided to describe the story, following the guidelines of the STEP test. Second, learners were told that they would do the same task again, but before that they were asked to do one of the three Planning Activities. Specific instructions were given for each activity. For the Writing activity, they were told to write out what they had wanted to say when they performed the Unplanned Task. Following many previous studies, the time allowed was 10 minutes. No detailed planning instructions (e.g. “try not to write out everything in detail” (Mehnert 1998:89) or “try not to write full sentences” (Ortega 1999:123)) were provided, since it was considered that restrictions such as those given in the previous studies would be ignored. As Ortega (1999) found, learners reported that they were “running through it a couple of times,” “talking to myself” and “rereading myself” (p. 127) while planning, which is presumably done at the sentence level as well as the phrase level. For Rehearsal, they were told to rehearse by saying aloud what they had tried to say in the Unplanned Task as often as they wanted until the 10-minute time limit was up. For Reading, they were provided with a model passage of the picture story (i.e., L2 input) and told to read it silently for the allotted 10 minutes and to think how they could redo the task.

When the learners had finished the assigned planning activity, they were asked to complete the questionnaire. The questionnaire consisted of 7 written questions about the degree of attention to language (e.g., vocabulary, grammar, structure, and pronunciation) as well as the usefulness of each planning activity. The learners were asked to choose on the scale of 1 (not at all) to 7 (very much so).

After the questionnaire was completed, the learners carried out the two-minute Planned Task. When producing both the Unplanned and Planned narratives, they were allowed to look at the picture set to which they were assigned. To encourage the use of past tense forms, an introductory sentence written in the past tense was provided for each of the picture sets, and learners were required to start with it when they told the story.

Measures

1. *Fluency*

a. Rate of speech

Rate of speech was operationalised as amount of speech. This measure was recommended by Esser (1995), who demonstrated that the quantity of speech served as an indicator of fluency. When the number of words was counted, fillers both in Japanese and English, such as “uh” “unn” and “well” were excluded. A repeated word and a self-corrected word were counted as separate words. Contractions, such as “wasn’t” and “he’s”, were counted as two words.

b. Repetitions

The second measure of fluency, frequency of repetitions, indicates how frequently learners use hesitations in constructing a sentence or to gain processing time (Bygate 1996; Foster & Skehan 1996; Ortega 1995; Skehan & Foster 1999; Riggensbach 1991). According to Bygate (1996), repetitions can be categorised as verbatim repetitions or substitutive repetitions. A verbatim repetition “occurs when hesitating, creating time to find an appropriate word”, while a substitutive repetition “seems to be employed when correcting a word or grammatical feature” (p. 141). Bygate used the frequency of these repetitions to indicate changes in the level of fluency between the first and second performances in his study of task repetition. In the current study, the percentage of repeated words per performance was used in the same way, since the amount of speech varied greatly between the Unplanned and the Planned Tasks. In the following examples, the underlined words are verbatim, and the italicised words are substitutive repetitions.

- (1) A girl said you should . be . you should be . *you should throw* can in a dust box.
- (2) Many people . many people drink drink and eat . *many people are eating and*.

2. *Complexity*

a. The number of clauses per T-unit

A T-unit is a measure of the linguistic complexity of sentences. It is defined as “consisting of one independent clause together with whatever dependent clauses are attached to it” (Richards, Platt, & Platt 1996: 390). The T-unit analysis was initially developed to assess written language and has been replaced by the c-unit analysis for oral language. However, the present study used the

T-unit rather than c-unit (Foster & Skehan 1996) on the grounds that a non-interactive narrative task involves few non-finite units. Since non-native oral production is not always complete and correct, I followed the procedure used by Brock (1986:52), that is, “a segment of speech was not disqualified as a T-unit because it lacked or included incorrectly the copula, prepositions, articles or inflectional morphology.”

I referred to *COBUILD English Grammar* (1992) for definitions of finite clause and non-finite clause. Repeated clauses, partly self-corrected clauses, or incomplete/abandoned clauses in the same T-unit were not counted. The following are some of the examples. The underlined clauses were excluded in counting.

- (1) On a top of it, they had to carry the rubbish . refusal refusal sack, . because they couldn't just leave that . they just couldn't leave it as it was.
- (2) Instead, he put . he threw a can without thinking anything.
- (3) And then, the girl was called Mr. Tanaka and said . and said and told him to put to the empty can in the bin.
- (4) And then, . . . a very pretty girl came to came near him . came to his table, and . . .

b. T-unit length

The number of words per T-unit has been employed as a measure of language complexity in writing (Cooper 1976; Hirano 1991). This measure can also be used to assess complexity in speaking (Kawauchi & Kamimoto 2000). Repeated words were excluded.

c. Subordinate clauses

Following Crookes (1989), Wigglesworth (1997), and Bygate (1996), this study also used the number of subordinate clauses. Crookes and Wigglesworth employed the frequency of subordinate clauses per utterance or T-unit. However, this study used the frequency of subordinate clauses per performance as in Bygate, since the time allocated for each performance was limited to two minutes and was the same for all the participants.

d. The number of word types

Finally, the number of different words was examined as a measure of lexical complexity. Different words are called “types” and show lexical variation. However, lexical variation in the form of the type/token ratio is often claimed to be affected by differences in text length (Laufer & Nation 1995). Instead of type/token ratio, the number of types per performance was used in this study.

It was assumed that the number of different words would be a good basis on which to assess the lexical richness of the learners' productions.

3. Accuracy

To measure accuracy, I used the past tense markers for copula *be*, auxiliary verbs, regular verbs, and irregular verbs following Ellis (1987) and Bygate (1996). Bygate (1996) reported an increase in the use of the regular past, as opposed to irregular past, in the repeated task. Ellis (1987) showed that significantly more correct uses of copula were produced in the planned narrative, or the narrative performed after writing, than in the unplanned narrative, or the narrative without prior writing. He also revealed that the use of irregular verbs did not differ across the three tasks (i.e., writing, writing then speaking, and speaking without writing).

According to Tono (2000), the copula and the auxiliary reached 90% accuracy, and even the past tense of irregular verbs reached about 84% accuracy in Japanese learners by the third year of senior high school. However, whereas Japanese learners may well have acquired an underlying knowledge of the past tense before entering university, they may not be able to make use of this knowledge during speaking. In fact, there is a lot of empirical evidence to show that these morphemes are often omitted or misformed in oral tasks (see the transcriptions in Williams 1999). It is, therefore, possible that the strategic planning will enable learners to attend to these morphemes and facilitate their correct use in subsequent performance.

To measure accuracy, I basically followed Ellis (1987) and examined suppliance in obligatory occasions. Any verb which could not be clearly transcribed, such as “noticed to” where assimilation of the /t/ sound occurred, was excluded from the analysis. Immediate repetitions of any verb, which is common in oral tasks, were not counted, either. In cases where the first attempt was erroneous and the second attempt correct, or vice versa, both attempts were included in the scoring. No allowance was made for a verb that was marked for past tense but marked incorrectly (e.g., standed, hitted). The uninflected verbs like “hit” and “put” were excluded when the use of the past tense was not obvious. The failure of agreement (e.g., there was only two children) was not counted as an error, but counted as correct, because the focus was on past tense formation. When the past perfect form was used (e.g., had annoyed), the auxiliary verb and past participle were scored separately. Copula in this study consisted of “was” and “were” and other auxiliary verbs like “could” and “did”. When the copula was omitted (e.g., “they in the mountain” and “they walking in the mountain”), it was judged to be erroneous.

Analysis

The oral data were transcribed by the present author and double-checked by a senior student assistant who had just returned from a study abroad program. As for complexity and accuracy, two raters examined a randomly selected sample of 30% of the total data for each measure. They were Japanese teachers of English (the author and her colleague) with more than 10 years teaching experience in a Japanese university. As for repetitions, intra-rater reliability was calculated. The author examined the data on two separate occasions (three weeks apart). The number of words and types were counted using WordSmith.

The results of inter-rater reliability for complexity indicated 91.8% agreement, while that for accuracy was 92.1%. In case of disagreement, a third rater, a native American teacher of English judged the data. With this high reliability, the remaining data were examined by myself on two different occasions (three weeks apart); there was a 97.4% intra-rater agreement. Concerning the assessment of repetitions, intra-rater reliability reached 92.8% agreement. These agreement rates can be considered sufficiently high (Hatch & Lazaraton 1991).

As explained above, this study used three different sets of pictures. One way to confirm if these picture tasks are equivalent is to examine how the participants performed under the Unplanned condition. In order to see whether or not there was a significant difference, all the data from the Unplanned performances were examined using one-way ANOVAs (3 picture sets) for fluency [F], complexity [C], and accuracy [A]. Table 2 shows the results. None of the results yielded any significant differences. It can be said, therefore, that the learners performed these three tasks in very similar ways when they were faced with each picture set for the first time.

Table 2. ANOVA results (F-values) for three picture sets in unplanned performances

Measures	Low EFL	High EFL	Advanced ESL
[F] Number of words	1.97ns	0.22ns	0.21ns
[F] Repetitions	1.06ns	0.61ns	0.56ns
[C] Clauses per T-unit	0.58ns	1.27ns	0.01ns
[C] Words per T-unit	0.2ns	2.42ns	0.85ns
[C] Subordination	1.21ns	3.20ns	0.12ns
[C] Number of types	0.54ns	0.25ns	0.02ns
[A] Correct past tense (%)	0.66ns	0.03ns	0.33ns

Note. ns = non significant at $p < .05$.

The L2 input used in the Reading activity measured 77.4 for Library, 86.2 for Jogging, and 74.7 for Hiking on the Flesch Reading Ease scale, suggesting that the stories were not difficult and did not vary greatly in readability .

All the statistical analyses in this study were carried out using the statistical package StatView, version 4.5.

Results

Table 3 presents a summary of the descriptive statistics for all the measures for the Low EFL, High EFL, and Advanced ESL groups. It shows the mean scores and standard deviations for fluency [F], complexity [C], and accuracy [A].

Research Question 1: What effect does proficiency have on L2 learners' performance of an oral narrative task after strategic planning?

Table 3. Descriptive statistics for fluency, complexity, accuracy of the three groups

Measures	Low EFL	High EFL	Advanced ESL
[F] Number of words			
Unplanned	71.33 (23.03)	113.14 (35.98)	138.27 (34.27)
Planned	97.79 (25.59)	142.64 (34.18)	154.88 (35.11)
[F] Total Repetitions (%)			
Unplanned	13.31 (10.25)	12.20 (8.76)	5.97 (4.29)
Planned	11.06 (7.31)	8.9 (8.19)	7.08 (5.79)
[C] Clauses per T-unit			
Unplanned	1.35 (0.23)	1.73 (0.37)	1.91 (0.38)
Planned	1.52 (0.27)	1.92 (0.38)	2.13 (0.39)
[C] Words per T-unit			
Unplanned	9.15 (2.33)	9.93 (2.27)	11.21 (1.93)
Planned	9.75 (2.01)	11.06 (2.17)	12.02 (1.64)
[C] Subordination			
Unplanned	0.77 (0.91)	2.75 (1.65)	4.82 (2.46)
Planned	1.85 (1.21)	4.36 (2.23)	5.21 (2.51)
[C] Number of types			
Unplanned	36.65 (9.48)	53.86 (12.82)	69.21 (17.29)
Planned	50.42 (11.38)	68.53 (12.69)	77.67 (13.74)
[A] Past tense correct (%)			
Unplanned	40.10 (23.06)	73.50 (20.89)	71.73 (20.01)
Planned	54.61 (25.29)	81.34 (16.06)	75.85 (17.77)

Key: () Standard deviation

In order to examine the first research question, the linguistic outcomes from the Unplanned and the Planned performances were compared. As can be seen in Table 3, almost all the measures favoured the Planned over the Unplanned. These results were submitted to two-way ANOVAs (3 proficiency levels \times 2 planning conditions) with repeated measures for each dependent variable (fluency, complexity, and accuracy scores). A two-way ANOVA is considered viable (rather than a multivariate ANOVA) since the structure of the dependent measures is clearly defined in terms of the three separate areas of fluency, complexity, and accuracy (cf. Skehan & Foster 1999). The Bonferroni correction was applied where *F* values justified this procedure. Table 4 presents the results.

First, we will consider the results for fluency. There were main effects for proficiency and planning on both fluency measures. There was also an interaction between planning and proficiency for the number of words, so a post-hoc analysis was carried out. This showed that in the Unplanned condition there were significant differences among the three proficiency groups, resulting in the order of Low < High < Advanced, but in the Planned condition no difference was found between the High EFL and Advanced ESL groups, i.e. the order was Low < High = Advanced.

As for total repetitions, the Bonferroni correction showed that the result for the Advanced ESL group was significantly different from those for the Low and High EFL groups, with the latter two groups showing no significant difference, making the order Advanced < Low = High. Although the Advanced ESL group used the fewest repetitions, their repetitions in the Planned Task were significantly more frequent than those in the Unplanned Task. As seen from Table 2, the Advanced ESL group showed an increase in the number of repetitions from the Unplanned Task (5.97%) to the Planned Task (7.08%). This suggests that where repetitions were concerned the strategic planning had a negative effect on fluency for this group.

The repetitions were then examined in more detail based on the types (verbatim or substitutive repetitions) by means of two-way ANOVAs. Verbatim repetitions displayed the same tendency as total repetitions shown above. On this occasion, however, the Advanced group showed a significant decrease in the number of verbatim repetitions: from 3.37% in the Unplanned Task to 2.49% in the Planned Task. The results for substitutive repetitions were revealing. The main effect for proficiency was significant ($F = 4.00, p < .021$), and this was due to the difference between the Advanced ESL and the Low EFL groups (Advanced < Low). Interestingly, there was no main effect for planning condition ($F = 0.52, p < .47$). Also, no interaction was found. It is clear from

Table 4. ANOVA results for fluency, complexity, and accuracy

Measures	SS	DF	MS	F	Sig. of F
[F] Number of words					
Proficiency	156010.73	2	78005.36	46.80	<.0001
Planning	33477.83	1	33477.83	135.54	<.0001
Proficiency × Planning	1738.01	2	869.00	3.52	.045
[F] Repetitions					
Proficiency	1275.77	2	637.89	7.07	.0013
Planning	154.86	1	154.86	4.58	.0345
Proficiency × Planning	182.98	2	91.49	2.71	.071ns
[C] Clauses per T-unit					
Proficiency	14.68	2	7.34	46.20	<.0001
Planning	2.16	1	2.16	33.77	<.0001
Proficiency × Planning	0.03	2	0.01	0.27	.764ns
[C] Words per T-unit					
Proficiency	184.57	2	92.29	14.31	<.0001
Planning	39.55	1	39.55	16.70	<.0001
Proficiency × Planning	2.88	2	1.44	0.61	.546ns
[C] Subordination					
Proficiency	528.42	2	264.21	55.02	<.0001
Planning	59.35	1	59.35	29.66	<.0001
Proficiency × Planning	13.56	2	6.78	3.89	.023
[C] Number of types					
Proficiency	34377.22	2	17188.61	62.88	<.0001
Planning	8623.44	1	8623.44	157.08	<.0001
Proficiency × Planning	428.67	2	214.33	3.90	.031
[A] Correct Past tense %					
Proficiency	40958.60	2	20479.30	26.99	<.0001
Planning	4441.42	1	4441.42	31.95	<.0001
Proficiency × Planning	1053.25	2	526.63	3.79	.033

these results that substitutive repetitions were used both in the Unplanned and Planned performances alike. Thus, although the Advanced ESL group employed significantly fewer substitutive repetitions than the Low EFL group, a close look at the results shows that the Advanced learners employed substitutive repetitions more frequently in the Planned Task (2.60% in the Unplanned and 4.59% in the Planned Task).

We will now consider the results for complexity. Main effects were found for proficiency and planning conditions with no interactions on two of the measures (i.e. the number of clauses per T-unit and the words per T-unit). The Bonferroni correction showed that the Low EFL group differed significantly

Table 5. Overview of accuracy based on verb types (% correct)

	Low EFL		High EFL		Advanced ESL	
	Unplanned	Planned	Unplanned	Planned	Unplanned	Planned
Past copula	64 (32.3)	113 (44.3)	125 (66.5)	147 (76.6)	134 (66.3)	151 (68.6)
Past regular	40 (36.7)	91 (52.6)	86 (73.5)	134 (76.1)	76 (65.6)	117 (72.7)
Past irregular	91 (53.2)	155 (68.3)	173 (76.2)	220 (84.9)	190 (80.9)	251 (87.5)

from the High EFL group, which also differed from the Advanced ESL group (Low < High < Advanced). The results for T-units also showed that the Planned performances were significantly more complex than Unplanned performances.

The results for subordination and word types yielded a significant interaction between the proficiency and the planning conditions. Further analyses revealed that, as for subordination, the Low EFL group differed significantly from the High EFL and the Advanced ESL groups, but no significant difference was found between the latter two groups under the Planned condition (Low < High = Advanced). Moreover, the Advanced ESL group did not differ significantly in the Planned and Unplanned Tasks. This indicates that strategic planning did not facilitate the use of subordination for Advanced ESL learners. The analysis of the interaction regarding the word types showed the order to be Low < High < Advanced for both the Unplanned and Planned Tasks.

Finally, the effects on accuracy were examined. Main effects for proficiency and planning conditions were found for correct past tense and there was also an interaction. In both the Unplanned and Planned performances the Low EFL group was significantly different from the Advanced ESL and High EFL groups, but no significant difference was found between the latter two groups (Low < High = Advanced). Similar to the results for subordination, the Advanced ESL group did not perform significantly differently in the Unplanned and the Planned Tasks.

How were the different verb categories (copula, regular verbs, and irregular verbs) used under the two planning conditions? To see whether or not there was a difference, the frequency and the percentage of correct use of each verb type were examined. Table 5 shows the results.

It is clear that the accuracy levels for past tense forms vary according to the verb categories. Accuracy in the use of past irregular verbs was the highest in both the Unplanned and Planned Tasks, while the past copula tended to be the least accurate. The difference in frequency of correct and deviant past tense forms for each verb category was analysed using chi-squared. As for past irregular verbs, all the groups yielded significant differences between the

Table 6. Gains in planned performances and ANOVA results

Measures	Low EFL	High EFL	Advanced ESL	F	Sig. of F
[F] Number of words	+25.33 (15.94)	+29.50 (21.94)	+16.61 (29.87)	2.96	.056ns
[F] Repetitions (%)	-2.56 (10.85)	-3.01 (5.48)	+1.11 (5.50)	2.61	.078ns
[C] Clauses per T-unit	+0.18 (0.30)	+1.83 (1.86)	+0.52 (2.08)	12.83	<.0001
[C] Words per T-unit	+0.85 (2.11)	+1.13 (1.96)	+0.91 (2.37)	0.19	.831ns
[C] Subordination	+1.08 (1.33)	+1.61 (2.06)	+0.39 (2.11)	3.89	.023
[C] Number of types	+13.87 (8.10)	+14.67 (10.73)	+9.18 (12.72)	2.84	.062ns
[A] Past tense (%)	+14.52 (18.56)	+7.56 (14.48)	+4.13 (15.65)	4.17	.018

Note. () standard deviation

Unplanned and the Planned Tasks: Low EFL ($\chi^2 = 9.38, p < .05$), High EFL ($\chi^2 = 5.96, p < .05$), and Advanced ESL ($\chi^2 = 4.30, p < .05$). The accuracy for past copula significantly improved in the Planned Task only for the Low EFL group ($\chi^2 = 6.73, p < .05$) and the High EFL group ($\chi^2 = 4.10, p < .05$). As for past regular verbs, only the Low EFL group showed a significant increase in accuracy ($\chi^2 = 6.80, p < .05$).

To investigate in more detail the effects of planning for the three groups, gains from the Unplanned to the Planned performances were examined. That is, the differences between the Unplanned and the Planned Tasks were compared for the three groups. By so doing, it will become clearer which proficiency group benefited most from the opportunity to plan. One-way ANOVAs (3 proficiency groups) were carried out separately for each measure. The results are shown in Table 6.

In the case of fluency, no significant differences were found in the gains among the three groups. However, the results for the High EFL group showed the highest increase in the number of words and greatest decrease in repetitions, suggesting that the High EFL learners benefited more than the other two groups.

Significant differences were evident in two of the complexity measures: the number of clauses per T-unit and subordination. Concerning the gains of clauses per T-unit, the Bonferroni correction revealed that the High EFL group was significantly different from both the Low EFL and Advanced ESL groups, but the latter two groups did not differ (High < Low = Advanced). As for subordination, the High EFL group showed the largest increase, which was significantly greater than the increase in the Advanced group (Advanced < High).

There were also differences among the three groups in the gains for accuracy. This time, the Low EFL group showed the greatest gain, which was

significantly different from that of the Advanced ESL group (Advanced < Low). In fact, the gain in accuracy in the Low EFL group is more than three times larger than that for the Advanced ESL group. However, although these gains demonstrate an effect for planning, the high standard deviations suggest that there was considerable intra-group variation.

Research Question 2: To what extent is the effect of proficiency on L2 learners' performance of an oral narrative dependent upon the type of strategic planning?

To examine the individual effects of the strategic planning activities, the results for the Planned performances, which have been combined up until now, were considered in relation to the three planning activities. Two-factor ANOVAs (3 proficiency levels \times 3 types of pre-task activity) with repeated measures were carried out. As might be expected from the preceding analyses, there were main effects for proficiency in all the measures except for repetitions. However, no differences in the effects of the three planning activities on fluency, complexity, and accuracy were found. No interactions were found, either. The lack of significant differences for the planning activities suggests that there are no distinctive effects pertaining to the three types of planning.

Discussion

This study investigated two research questions relating to the roles of proficiency and the type of strategic planning. I will first summarise the results and then discuss the main findings.

The results show that strategic planning had beneficial effects on the fluency, complexity, and accuracy of these Japanese learners' oral narratives. Enhanced fluency in the Planned performance was evident in both the greater number of words produced and in the reduced number of repetitions. However, planning seems to have benefited the High ESL in particular as it enabled them to perform as fluently as the Advanced ESL learners in the Planned performance whereas they were notably less fluent in the Unplanned Task. Also, although, the Advanced ESL learners produced the fewest repetitions overall among the three groups, they employed significantly more repetitions, particularly substitutive repetitions, in their Planned performance than in their Unplanned performance, indicating that where this aspect of fluency was concerned the planning had a negative effect for learners of this level of proficiency.

Greater complexity was also evident in the Planned performances of all the groups. However, the Advanced ESL group failed to show greater complexity in

the case of subordinate clauses in the Planned Task. Planning seems less helpful to this group than the High EFL group, which performed the Planned Task at a similar level of complexity to the Advanced ESL group.

The Advanced ESL group also failed to show any significant increase in accuracy of past tense verb forms in the Planned Task suggesting that the planning effect on accuracy was very limited for this group. The accuracy results for the Advanced ESL group were similar to those of the High EFL group in both the Unplanned and Planned Tasks. Concerning the three types of past tense forms, the present study reported results that were somewhat different from those of Ellis (1987) and Bygate (1996). The accuracy rate for irregular past was highest in the Unplanned Task for all the groups and increased significantly for all groups in the Planned Task. In Ellis (1987), no significant improvement for this structure was observed in the planned condition. Bygate (1996) reported an increase in the accuracy of use of regular past, but the present study found that only the Low EFL group benefited significantly from the opportunity to plan. This may reflect the fact that the Low EFL group's score for regular past tense in the Unplanned Task was very low and thus this group had the greatest room for improvement.

As might be expected, the performance of the three groups in the Unplanned Task corresponded to their proficiency level. That is, overall, the Low EFL Group was the least fluent, used the least complex language and was the least accurate while the Advanced ESL Group performed the best with the High EFL Group intermediate. A key finding of the study was that for some measures the High EFL group was able to catch up with the Advanced ESL group in the Planned Task. In other words, the Advanced ESL group did not always produce the best performance in the Planned Task.

This was also evident in the comparison of gains by the three groups. The gains for fluency did not show any significant differences among the three groups, but gains were greatest for the High EFL in all the measures. In the case of the gains for complexity, the High EFL group surpassed the Advanced ESL group in clauses per T-unit and in subordination. These findings indicate that learners in the High EFL group benefited the most from the opportunity to repeat the task after planning. Where accuracy was concerned, however, it was the Low EFL group that manifested the largest gains, which were significantly greater than the gains for the Advanced ESL group.

Thus, it can be concluded that the High EFL group benefited most from the opportunity to plan in the case of fluency and complexity, while the Low EFL group did so in accuracy. The Advanced ESL group appeared to benefit much less than the other two groups. These findings are not accordance with those

of Wigglesworth (1997), who found a planning effect for accuracy mainly in her higher proficiency learners. The present results also fail to support Ortega's suggestion (1999) that, in the case of accuracy, planning might have the greatest effect on advanced level learners.

Why does planning seem to affect each group differently? In the case of the Advanced ESL learners there might have been a ceiling effect. They seem to be able to handle the task well even when it is unplanned, possibly because their L2 knowledge is sufficiently proceduralized to enable them to access it easily in rapid on-line planning. In contrast, the Low EFL learners appear to be focusing only on meaning and attending least to form in the Unplanned Task, as is suggested by VanPatten's (1996) study and the findings of Skehan and Foster (1999). Planning provided them with the opportunity to attend to form when they repeated the task. However, in contrast to the High EFL group they had limited L2 knowledge to draw on in the first place. Thus, the High EFL benefited most from the planning because (1) they had adequate L2 knowledge to draw on and (2) the planning enabled them to access this knowledge when they repeated the task.

The findings for the second research question indicated that the different types of planning did not influence the learners' performance of the Planned Task, despite the fact that there are clear differences in the resources available when learners were engaged in the different planning activities. Nor did the learners' level of proficiency have any effect on their ability to benefit from the different types of planning. It seems that whatever the type of planning and whatever the level of proficiency, some improvements in performance follow when the opportunity to plan is available.

However an inspection of the transcripts of the learners' performance of the tasks does suggest that there may be some differences related to the planning activities. Differences were evident in the learners' use of some low frequency lexical items (e.g., *vending machine*, *empty*, and *librarian*) and problematic structural items (e.g., *play with*, *throw OBJECT over one's shoulder*, and *take OBJECT home*). In the Unplanned performance, these items were not used at all or they were paraphrased in a simple way, suggesting that learners did not know them. However, those who engaged in the Reading activity tended to benefit from the L2 input available to them, showing more target-like use of these items in the Planned Task. The following examples illustrate how learners changed their use of language. The difficult lexical and structural items are underlined.

Noh: Low EFL: Unplanned: Jogging

There are two chi children . they . play . ball . When Mr. Tanaka was jogging . He ... *mikke (find)* . . see juice box . He want to drink juice . So it was very hot day . Mr. Tanaka bu . bought a juice and drink . So . Mr. Tanaka . finished drink . . he throw away a . . juice cans . . .

Noh: Low EFL: Planned (Reading): Jogging

He . he saw a vending machine eh . eh . There are children . The boy and the girl were playing with a ball near the vending machine so when he nn . when he saw the vending machine he want to drink . Then he decide to . decided to buy . . ah he decided to stop and buy a soft . cool drink . He . bu . bought soft cool drink and he drank it . nn . The girl was . uh the girl standed near Mr. Tanaka . When he finished . . drank finished nn . finished his drink . he threw the empty can . . .

In contrast, those learners who completed the Writing and Rehearsal planning activities did not use these difficult items. Instead, they embellished their stories with their own “interpretations” (Tannen 1984:33) or “evaluative comments” (Bygate 1996:141). These were often found in the introductory part of the story. Some examples are provided below:

Yam: Low EFL: Planned (Writing): Jogging

The park is Ohori Park in Fukuoka city . Mr. Tanaka is a student of Kurume University . Usually he .. every . every Sunday he jogs . jogs in the park because he . . he doesn't . he doesn't . do any exercise . everyday . so ... he jogs every Sundays . . .

Chika: High EFL: Planned (Rehearsal): Hiking

On weekdays they have work each other . they were working as a teacher and they were looking forward holiday . As they expected . it was very fine . sunny day . so at first they were really enjoying the walking . because they could find so many wild flowers and birds and they could enjoy the beautiful scenery . . .

What these examples suggest is that learners may have a different focus according to the planning activity. After engaging in Reading, their attention seems to focus on specific linguistic problems, which they solve by referring to the L2 input. However, when engaged in Writing and Rehearsal, learners appear to pay attention to other aspects of the task by adding interpretations and evaluative comments rather than dealing with linguistic problems. The Low EFL learners, in particular, seem to have benefited from the Reading planning activity.

The responses to the questionnaire go some way to confirm this. The Low EFL learners indicated that they paid the greatest attention to vocabulary when

engaged in Reading, which was significantly different from Writing and Rehearsal (Friedman $\chi^2 = 6.2, p < .05$). In fact, they gave significantly higher ratings to Reading in most of the questions. In contrast, High EFL and Advanced ESL learners responded in a very similar way to all three planning activities.

Conclusion

Despite a number of limitations (e.g. small sample size; only one type of task was investigated), this study provides clear evidence that learners' L2 proficiency is a factor in determining whether and to what extent strategic planning affects subsequent task performance. Thus, the High EFL learners tended to benefit the most in the case of fluency and complexity while the Low EFL learners appeared to gain most in accuracy. The Advanced ESL learners gained the least. Thus, whereas the Advanced ESL learners performed the best in the Unplanned Task (as might be expected), their performances in the Planned Task tended to be equalled by the High EFL learners. The results indicate that there may be a level beyond which planning will have only a limited effect.

This study also investigated the effects on task performance of different types of planning. However, the results did not reveal any clear differences as far as the quantitative analysis was concerned. In contrast, a qualitative look at the language use evident in the Planned Task suggested that there were some differences in task performance resulting from the different types of planning. In particular, there appeared to be differences reflecting whether the planning was input-based (Reading) or output-based (Writing and Rehearsal). These qualitative aspects of planning are in need of further study.

SECTION IV

Within-task planning

Within-task planning was defined in Chapter 1 as the on-line planning that takes place during a task performance. When this is unpressured the participants have the opportunity to conceptualise, formulate and articulate their messages with some care. The two chapters in this section examine within-task planning, albeit in very different ways.

Ellis and Yuan operationalize within-task planning experimentally in terms of the performance conditions for competing the task (another narrative task). Whereas two groups (one performing an oral and the other a written narrative) were pressured to perform the task rapidly, another two were allowed to perform it in their own time. The problem with this approach, as Skehan and Foster point out, is that it does not demonstrate that there are actual processing differences in the way the pressured and unpressured groups performed the task. They suggest that it might be possible to distinguish rapid and careful on-line processing in terms of a cluster of variables associated with repair fluency (i.e. mid-clause pauses, filled pauses, reformulations, repetitions and false starts). These features are also associated with shorter length-of-runs. They constitute psycholinguistic evidence that learners are engaging in on-line planning. Interestingly, Skehan and Foster found differences in some of these features between the first five minutes and second five minutes of their participants' performance of a decision-making task. One explanation of these differences is that, as the task progressed, the pressure on on-line planning built up with a consequent effect on fluency.

Both chapters build on previously published studies. Ellis and Yuan's study constitutes an attempt to disentangle the effects of modality (writing vs. speaking) and on-line planning (pressured vs. unpressured) that were confounded in Ellis (1987). Ellis and Yuan were able to show that both variables have an impact on complexity and accuracy, with the impact of modality being much the stronger. Skehan and Foster's study is a partial replication of Foster and Skehan (1996) where the different effects on task performance of detailed and undetailed strategic planning were examined. The earlier study indicated that it was the undetailed planning condition that promoted greater accuracy but

this result was not repeated here. In this study it was the detailed planners who were more accurate (although not statistically significantly so).

Theoretically, both studies in this section are of interest. Ellis and Yuan's study demonstrates that planning effects are evident in written as well as oral performance. Theirs is the only study in the book that considers written task-performance. Foster and Skehan's study is important because it provides convincing evidence that the effects of planning may not be consistent throughout the duration of a task. They found clear differences between the strategic planning and the no-planning groups in the first five minutes of the performance of the task but less clear evidence of any differences in the next five minutes. This study suggests the importance of distinguishing between learners' use of language early on and later on in a task.

There are also a number of pedagogic implications of the results from these studies. Ellis and Yuan's study suggests that the on-line conditions under which learners are asked to produce both oral and written texts will influence the sophistication and accuracy of the language they produce. Thus, their study reinforces the widely held view that L2 learners will have difficulty in accessing their full linguistic competence under examination-type conditions when they are pressured to speak or write rapidly. Skehan and Foster's study indicates that, although, as previous studies have shown, providing learners with the opportunity to plan strategically will aid their performance of a task, enabling them to attend to form, such opportunity may not have a prolonged effect. Thus, if the aim to use strategic planning to assist learners to achieve high levels of complexity and accuracy, teachers might do better to employ tasks which can be completed in a relatively short period of time. More generally, both studies suggest ways in which the effects of on-line planning can be manipulated to both favour 'best' performances and to prepare learners for real-life situations.

CHAPTER 6

The effects of careful within-task planning on oral and written task performance

Rod Ellis and Fangyuan Yuan

University of Auckland / University of Pennsylvania

Introduction

This chapter differs from the other chapters in the book in two principal ways. First, it focuses on within-task planning rather than strategic planning. Second, it examines the effects of planning on written as well as oral performance.

As we saw in Chapter 1, within-task planning refers to the planning that takes place on-line, during as opposed to before the performance of the task. Within-task planning can be ‘careful’ in the sense that performers of the task have ample opportunity to plan their productions and make use of this opportunity to attend to the content and/or expression of their performance. Alternatively, it can be ‘pressured’ in the sense that performers are required to produce text rapidly and thus have limited opportunity to attend closely to content and/or expression as they perform the task. Because short-term memory is of limited capacity (Baddeley 1986), the extent to which on-line planning is pressured will influence the nature of the planning processes that take place during performance. When pressured, performers are likely to need to prioritise some planning processes over others. In the case of careful within-task planning, however, they will be better able to attend to the full range of processes, including those that are more demanding on working memory. For these reasons, the nature of the on-line planning is hypothesized to affect the quantity and quality of the texts produced.

As we will see the planning processes involved in speaking and writing are very similar (Kellog 1996). However, there are also differences. In particular, writing, by its very nature, provides greater opportunity for careful within-task planning because writers have more time for text production and thereby

greater control over the processes involved. Thus, it remains an empirical question as to whether the on-line planning processes involved in speaking and writing result in similar or different textual outcomes. It is this question that this chapter seeks to address.

An early study

The present study was motivated in part by an earlier study. Ellis (1987) investigated L2 learners' performance of written and oral narratives. In one condition they were given a picture composition and asked to write the story. They were given ample time (approximately one hour) to complete this task but were asked to start writing straight away. Thus, this condition can be characterized as 'writing/careful within-task planning'. In another condition, the learners were given a different picture composition and asked to tell the story orally after just two minutes to prepare it. This task might be characterized as 'speaking/pressured within-task planning' [1]. The written and oral narratives were scored for accurate use of three English past tense forms (regular past, irregular past and past copula). Table 1 presents the results. Clearly, all three past tense forms were performed more accurately in the written task than the oral task.

However, as has been pointed out by Crookes (1989), these results are not easy to interpret as the study conflates planning conditions and modality. That is, it is not possible to tell whether the greater accuracy evident in the written task reflects the fact that it involved writing or the opportunity for careful on-line planning. The purpose of the study reported below is to disentangle these two variables.

Also, Ellis' study only investigated accuracy, narrowly measured in terms of past tense forms. As we have seen in the earlier chapters reporting studies of strategic planning, subsequent research has incorporated criterion measures relating to broader measures of accuracy (e.g. percentage of error-free clauses) and measures of two other aspects of performance – fluency and complexity. In line with this research, the study reported here will also include measures of a wider range of textual variables.

Table 1. Accuracy of three past tense morphemes on two tasks (% correct)

Task	Regular Past	Irregular Past	Past Copula
Written	77	60	76
Oral	43	55	60

Other studies of on-line planning

Whereas there are a number of studies that have investigated the effects of strategic planning on oral task performance (e.g. Crookes 1989; Foster & Skehan 1996; Ortega 1999; Wendel 1997; see also the chapters in Section 3 of this book) there has been surprisingly little attention paid to within-task planning. Indeed, the studies of strategic planning listed above made no attempt to control for on-line planning; that is, they did not regulate whether the participants engaged in pressured or careful within-task planning. This may be one reason why these studies have produced varying results regarding the effect of strategic planning on linguistic accuracy, as the extent to which learners' productions are accurate may well depend on whether or not they engage in careful on-line planning rather than have opportunity for strategic planning (Wendel 1997; Yuan & Ellis 2003).

Variability studies are indicative of the kind of effect that careful planning can have on accuracy. Tarone (1982) proposes that L2 learners possess a continuum of styles, ranging from the 'careful' to the 'vernacular' (see Chapter 1). The former is operationalized in terms of whether the learner has time to attend to form, while the latter becomes evident in spontaneous natural language use, when learners are typically required to perform under pressure. This distinction corresponds closely to Ochs's (1979) 'planned' and 'unplanned language use. Ochs argues that, in the latter, learners have the opportunity to search their linguistic resources for grammatical information whereas in the former the lack of processing-time leads to speakers prioritising the search for lexical resources. From this point of view, then, it can be hypothesized that careful/planned speech is likely to be more accurate than vernacular/unplanned speech but not necessarily more lexically rich. While a number of studies (e.g. Dickerson 1975) have lent support to the hypothesis that careful speech is more accurate, other studies have shown that it is not quite so simple. Tarone (1985) has shown that whereas some grammatical structures (e.g. 3rd person -s) are performed more accurately in a careful style, other structures (e.g. articles and direct object pronouns) are more target-like in the vernacular style. A likely explanation for this is that the functional demands of a task may cause learners to attend to specific features even though these are difficult to process. Of course, these studies compared performance in different tasks; it is possible that if learners were asked to perform the same or a similar task under pressured and careful conditions, they would manifest accuracy in line with our hypothesis. Other, more recent studies (e.g. Bayley 1996; Regan 1996) demonstrate the ubiquity of 'style-shifting', dependent on learners' opportunity to plan care-

fully as they perform a task [2]. It is likely, however, that the degree of variation will still vary from structure to structure, depending, in particular, on whether learners have access to a simple and portable explicit rule (e.g. the rule for 3rd person -s). Ellis' (1987) study referred to above, for example, found a greater effect on regular past tense than on irregular past tense.

Another study of L2 variability helps to pinpoint the mechanism responsible for the linguistic variation that results from within-task planning. Hulstijn and Hulstijn (1984) investigated the effects of time pressure, focus of attention (i.e. whether on form or meaning) and metalingual knowledge on the accuracy of two Dutch word order rules in an oral story-retelling task. Interestingly, neither time pressure nor metalingual knowledge had any effect by themselves, whereas focusing attention on form increased the accuracy of both structures. This study suggests that careful within-task planning only enhances linguistic accuracy if it is used to attend to form. If it is used for some other purpose (e.g. to attend to propositional content of the message) no increase in accuracy occurs.

The effects of careful within-task planning on other aspects of oral L2 production have not been systematically investigated. In the case of fluency, increasing the opportunity for within-task planning can be hypothesized to differentially affect temporal aspects of production and hesitation phenomena. Thus, when there is opportunity to plan carefully, learners are likely to speak more slowly (e.g. pause longer and produce fewer syllables per minute) but may become less disfluent (e.g. make fewer filled pauses, repetitions and corrections) because they have more time to access their linguistic resources, including those that are not yet fully automatised. For the same reason, it also seems likely that increasing planning time will enhance the complexity of learners' productions (e.g. result in more subordination).

The within-task planning processes involved in writing have been more extensively researched, using think-aloud tasks. Hayes and Gradwohl Nash (1996), in a survey of planning research, note that planning and action are often interweaved in writing. This serves to both provide feedback on the effectiveness of the planning and to overcome memory limitations. They report that writers typically plan no more than 6–10 words before writing them down. In evaluating studies that have investigated the effects of planning on the quality of text production, they conclude 'text quality is strongly and positively related to time-on-task' (p. 53). In other words, the greater opportunity there is to plan carefully on-line, the better the written product.

Whereas Hayes and Gradwohl Nash considered only studies of L1 writing, other researchers have examined within-task planning in L2 writing. In a study

very relevant to our own research, de Larios, Marin & Murphy (2001) point out the importance of studying the allocation of 'processing time' when L2 learners write. They suggest that 'writers will differ in the way they adapt their time allocation to formulation processes as a result of task conditions' (p. 503). Using think-aloud protocols, they analysed the time that Spanish writers of L2 English allocated to three aspects of the writing process; planning (i.e. the retrieval and/or development of ideas), formulation (i.e. the production of pre-texts and texts) and revision (i.e. the changes made to written text). They found that, in general, the writers prioritised formulation but those with greater L2 proficiency were more likely to share composing time with the other processes. This study suggests that when L2 writers' working memories are under pressure (as is the case when proficiency is limited) they will concentrate on getting the message down but when their working memories are less taxed they will be able to attend to other aspects of the composing process. As Kellog (1996) notes, revision in particular makes heavy demands on the central executive of working memory with the result that it will be neglected when formulation demands are heavy. Whereas de Larios et al examined the role of L2 proficiency in on-line processing, the study below will consider time-on-task. We anticipate that this will have a similar effect (i.e. giving learners time to plan on-line will enable them to maximise their L2 proficiency).

Other studies of L2 writing have focussed on what aspects of text construction L2 writers attend to during composing. Whalen and Menard (1995) found that when writers were engaged in writing in their L2 they were more likely to attend to the linguistic level than in their L1 writing. They concluded that 'linguistic processing in L2 writing apparently inhibits more global processing at the textual and pragmatic levels' (p. 391). In other words, in comparison to L1 writers, L2 writers are more likely to focus on form to the detriment of organization and content.

These (and many other) studies provide some valuable insights into the nature of the on-line planning processes involved in L2 writing that are suggestive of how the opportunity to plan carefully on-line might influence output. However, they do not shed direct light on the effects of planning time on the quality of the texts produced. Furthermore, process-product studies of writing have relied on holistic or analytic ratings of text quality rather than the more precise discourse measures used by task-based researchers such as Skehan (1998a) and Robinson (2001c). In the study reported below we seek to examine how time-on-task effects the quality of written productions when this is measured in terms of fluency, complexity and accuracy. First, however, we will briefly examine the planning processes involved in speaking and writing

from a theoretical perspective as a basis for hypothesizing how time-on-task might affect production differentially in the two modalities.

Modelling the planning processes in speaking and writing

1. *Speaking*

The most influential model of speaking in studies of planning is Levelt's (1989) speech production model. As this has been described in some detail in previous chapters (see in particular Chapter 1), we will restrict ourselves to a broad outline of the model here, highlighting those aspects of it that we see as important for careful within-task planning.

The model identifies three key production processes. The conceptualizer establishes the intentions (goals) of the speaker and selects and orders the relevant information for achieving them. Its output is a 'pre-verbal message' (i.e. it is propositional in nature, not linguistic). This process involves both macro-planning, where the focus is on the goals, and micro-planning, which concerns the selection and organisation of the content needed to realise the goals. The formulator converts the pre-verbal message into a speech plan. This is achieved by identifying lemmas in the speaker's mental lexicon that match the propositional content of the pre-verbal message. The speaker selects lemmas for their meaning, which in turn activates lexical form. Lemma selection also provides the learner with the syntactic information needed to construct the surface structure of the message. While this is taking place morpho-phonological information relating to the lemmas is also activated and incorporated into the surface structure. The outcome of this process is a phonetic plan. The articulator then converts the phonetic plan into actual speech. Finally, there is a monitor. This works when the phonetic plan (from the formulator) and actual speech (from the articulator) is fed into the speech-comprehension system, which interacts with the production process to identify any mistakes that may have arisen. The model allows for parallel processing such that the output from one process can be fed into another process even if this is incomplete. That is, the processes occur simultaneously and dynamically.

As we noted in Chapter 1, Levelt's model was developed to account for speech production in the first language (L1) and may need to be adapted to account for speaking in an L2. De Bot (1992) suggests that in the case of the conceptualizer, macro-planning is not language specific but micro-planning is (i.e. the pre-verbal message specifies which language (or languages) are to be used to encode the message). De Bot argues that there are separate systems for the L1 and L2 as far as the processing components of the formulator are con-

cerned although the two systems are likely to be connected in at least some areas. In contrast, given the cross-linguistic influences evident in L2 pronunciation, he considers the existence of two separate systems for articulation ‘very improbable’ (p. 17). We might also note that whereas L1 speakers are generally able to carry out the processes involved in formulation and articulation (but not conceptualisation) without attention, L2 learners (especially those with limited L2 proficiency) are more likely to need to activate and execute their linguistic knowledge through controlled processing (McLaughlin 1987). Thus, they are likely to experience problems during the formulation and articulation stages, as these processes are demanding on working memory.

What differences can we expect in the operation of such a model depending on whether there is time for careful within-task planning or not? In careful within-task planning we would expect to find attention paid to all aspects of processing – conceptualising, formulation, articulation and monitoring. In pressured within-task planning, in contrast, we would expect attention to be paid to the three central processes (conceptualisation, formulation and articulation) as these are necessary for the production of speech but learners may have inadequate time to process the phonetic plan and to articulate speech through the speech-comprehension system, thus limiting their ability to monitor. We also suggest that limitations in working memory may inhibit the formulation process when speech is pressured. In such a situation, while learners may still be able to activate the lemmas and associated syntactic information needed to construct a surface structure for the message, they may lack the time they need to access the required morpho-phonological information from their lexicons. If this thinking is right, we would expect careful planning to result in speech production that is more accurate.

We also hypothesize that careful within-task planning will promote complexity as learners may use the time at their disposal to engage in more extensive micro-planning of the pre-verbal message. In this respect, careful within-task planning may work similarly to strategic planning. Finally, learners’ production is likely to become less fluent if they spend time on controlled processing during formulation and monitoring. In short, we anticipate that differences between careful and pressured within-task processing will be evident in all three aspects of language production.

2. *Writing*

As we noted in Chapter 1, there is no single model for writing that has the same status as Levelt’s model of speech production. The model that best suits our purposes here is Kellog’s (1996) model. This model explicitly relates processing

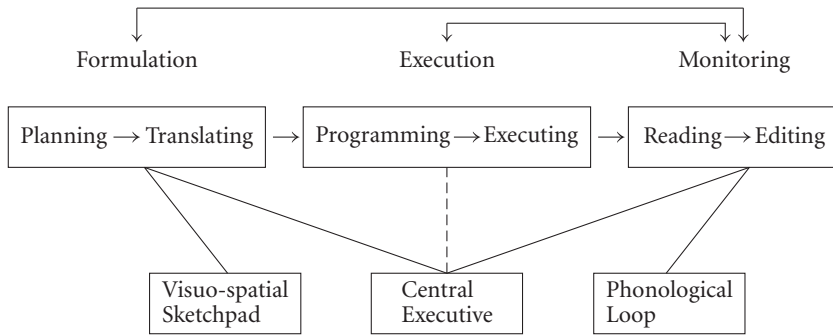


Figure 1. Kellogg's model of writing processes (Kellogg 1996:59)

components to Baddeley's theory of working memory. The model, which was described in some detail in Chapter 1, is shown diagrammatically in Figure 1. It distinguishes four basic systems involved in written text production ('formulation', 'translation', 'execution' and 'monitoring') with each system involving two principal components or processes. It should be noted that Kellogg's terms do not entirely match those of Levelt. Thus, for example, 'formulation', which involves 'planning' and 'translating', incorporates elements of Levelt's 'conceptualisation' and 'formulation'. As we also saw in Chapter 1, Kellogg suggests how the different components of the model relate to different components of working memory, emphasising the role played by the limited capacity of the central executive in the writer to make decisions about which writing process to prioritise when under pressure to produce text rapidly. Kellogg suggests that formulation demands are critical and will take priority over execution and monitoring when trade-offs are needed.

The similarity between Levelt's model of speech production and Kellogg's model of writing is striking. Both models posit an initial process where the goals and content of the message are established. Both incorporate a stage where the preverbal message is first lexicalised and then given a surface structure. Both include an articulation stage. And both recognize that language production can involve monitoring of the pre-production message and the articulated message. Furthermore, both models emphasise the continuous, overlapping nature of the processes involved. In what respects, then, are the processes of speaking and writing different? Arguably the differences are quantitative rather than qualitative in nature. That is, the pressure exerted on working memory is likely to be greater in speaking than in writing for, whereas the former occurs in real-time and is generally intolerant of significant pauses, the

latter, even when it involves freewriting, provides opportunities for the writer to take time-out from on-line production. In other words, the distinction between pressured and careful production in the case of writing is less clear-cut than in the case of speaking. There is another difference, also more qualitative than quantitative. Writing results in a visual object (the written message), which is amenable to inspection in a way that the aural trace left by a spoken message is not. It is likely that this facilitates the process of monitoring as it makes mistakes more salient. In other words, written text can be said to mediate self-correction more readily than oral text. In general, however, we would expect the same differences evident in pressured and careful oral production to emerge in written production, but, perhaps, to a lesser extent. That is we predict that, irrespective of modality, careful on-line planning will favour accuracy (and possibly complexity) at the expense of fluency.

L2 proficiency, in particular the extent to which learners possess the procedural knowledge needed for rapid on-line performance (Hulstijn 2002), is also a factor in both speaking and writing. Learners with limited procedural knowledge will experience difficulties in one or more of the component processes of speaking and writing. These difficulties will vary depending on what aspects of their knowledge are proceduralized. Bourdin and Fayol (1994), for example, found that L1 children's performance was better in speaking than writing, a finding they explained by the children's problems with graphic execution, which interfered with their ability to plan and formulate. Some adult immigrant learners may function similarly. In the case of adult L2 foreign language learners, however, the difficulties are more likely to lie in their lack of procedural knowledge of vocabulary and grammar. Such learners are likely to perform better in careful as opposed to pressured production and better in writing than in speaking because they will have more time to access their linguistic resources by means of controlled processes.

Research questions

1. Do L2 learners produce more fluent, complex and accurate oral and written language when they can plan carefully on-line than when they are pressured to speak/ write rapidly?
2. What effect does the modality of learner productions (oral vs. written) have on the fluency, complexity and accuracy of their language output?
3. Is there an interaction between the within-task planning condition and the modality of the learner productions?

Method

Design

The study involved a dual-factor (speaking and writing) between-participants design with two levels of within-task planning conditions (pressured planning and careful planning). The participants (Chinese learners of English) completed two tasks consisting of different sets of related pictures. One task required them to produce an oral narrative and the other task a written narrative. A total of 42 participants took part in the study. They were randomly assigned to one of three groups which were then asked to perform the two tasks in either the Pressured Planning or the Careful Planning condition as shown in Table 2 [3]. All the participants were administered a pre-test to ensure that the groups had equivalent English proficiency at the outset of the study. The textual products of the two tasks from both conditions were analysed in terms of fluency, complexity and accuracy.

Participants

The participants in the study were full-time undergraduate students who were English majors in the International Business Department of a Chinese university (i.e. foreign language learners of English). They were between 18 and 20 years old. At the time the data were collected, most of these learners had been learning English as a foreign language in Chinese schools for 8 years, first at elementary school and middle school and then in college. None of them had ever been to an English-speaking country and they had had little opportunity to use English for communicative purposes outside the classroom. Their scores in their Higher Education Bureau Examination [4] were between 100 and 120 (maximum possible = 150), with grades between A and B+ in the oral component of this examination. The participants can be considered to constitute a fairly homogeneous group of students in terms of their learning history and English proficiency.

Table 2. Design of the study

	Group 1 (N=14)	Group 2 (N=14)	Group 3 (N=14)
Oral Task	Pressured Planning	Careful Planning	
Writing Task	Careful Planning		Pressured planning

As college students, they had six hours of English each week, four hours for reading and writing and two hours for listening and speaking. Every two weeks, they had a one-hour oral English class from a native speaker of English from Canada.

All the students in two first-year classes were invited and agreed to participate in the study. In fact, they responded enthusiastically to the opportunity. They were told that the test and tasks they would complete were for purposes of research only and given that their teachers were not involved in the data collection in any way it seems likely that they accepted this at its face-value. They were not told the precise purpose of the research and they were assured that the information collected would not be used towards their course grades.

Pre-test material

The pre-test material was a version of the TOEFL (i.e. Test 1 from *Reading for TOEFL Workbook* published by the Educational Testing Service). The total test scores and the scores of the listening section were calculated and entered into one-way ANOVAs with the alpha set at .05. The listening section scores were examined separately on the grounds that they provide an indicator of the participants' on-line processing ability as the listening tasks required learners to process language in real time. The results of the ANOVAs revealed no significant differences across the three treatment groups in either overall TOEFL scores ($F = .39$; $p = .95$) or listening scores ($F = .464$; $p = .63$). Thus, it can be concluded that the three groups were equivalent in their English proficiency. Descriptive statistics are shown in Table 3.

Tasks

Narrative tasks were chosen to permit comparison with the results of studies that have investigated the effects of planning on similar tasks (e.g. Ellis 1987; Skehan & Foster 1999; Wendel 1997).

Table 3. Descriptive statistics for TOEFL scores across groups

	Group 1		Group 2		Group 3	
	Total	Listening	Total	Listening	Total	Listening
M	446.71	42.79	447.78	42.36	460.86	43.64
SD	35.34	3.68	27.84	3.65	26.57	3.46

Notes: M = mean SD = standard deviation

The tasks required participants to tell or write a story based on two different sets of six pictures taken from Heaton (1975). The story for the oral task was about three boys who could not get on a bus because four bigger boys pushed in front of them. They were forced to wait for another bus. However, their bus passed the first bus, which had broken down on a hill. The three boys laughed at the four bigger boys as they passed them. The story for the writing task was about a boy who got off a bus when it was dark and dropped one of the packages he was carrying. The boy set off home without noticing. A man picked up the package and chased after him to return it. The boy saw the man following him and was frightened. He began to run but eventually the man caught up with the boy and returned the package to him. This was the same picture story that Ellis (1987) used.

The task instructions were given in Chinese. All the participants were given the same prompts to establish the narrative genre required by the task. For the oral task the prompt was 'This morning, Tom, Jack and George ...' while for the writing task it was 'This afternoon, Tom ...'.

See Appendix A for the two tasks. While it is possible that the two tasks varied in complexity and thus posed differential encoding problems for the learners we consider this unlikely. Ellis (2003) identifies a number of criteria for evaluating the complexity of tasks. According to these, the two tasks can be seen as equivalent. They were both based on pictorial information, they contained a similar amount of information (6 pictures each), the information in both was dynamic and the narrative outcomes equally difficult to predict, there was a similar number of elements in the two stories, both tasks were context-dependent (the participants had access to the pictures as they told the stories) and the situations depicted in the pictures can be considered equally familiar to the learners. Furthermore, both tasks required a monologic performance and involved the same discourse mode (narrative). Similarly, an analysis of the tasks in terms of Robinson's (2001b) criteria for evaluating task complexity suggests the two tasks were equivalent. That is both picture series involved a similar number of elements, the tasks were of the here-and-now kind (as the participants had access to the pictures while they performed their narratives), the reasoning demands were roughly equivalent, both tasks were of the 'single task' (as opposed to 'dual task') type and the participants had no prior knowledge of the contents of either story.

Task conditions

In this study, within-task planning was operationalized at two levels: (a) Pressured Planning (PP) and Careful Planning (CP). The participants performed the task in their normal classroom setting. Both their normal teacher and the researcher were present.

1. *Pressured Planning (PP)*

a. Oral task

In this condition, participants were required to perform the task immediately after studying the pictures for a very short time (0.5 minute) and had to complete the task within a limited time (5 minutes). Thus, they had almost no time for planning the task in advance and were also pressured to perform the task rapidly so that opportunities for on-line planning were limited. To further increase a sense of pressure on the participants, they were required to produce at least 4 sentences for each of the 6 pictures.

b. Writing task

Participants were required to start the task immediately and to finish it within 17 minutes. They were asked to write at least 200 words. This was intended to limit the amount of time for extensive within-task planning while ensuring that it was possible for the participants to complete the story. A pilot study involving similar participants had been carried out to establish the time to be allowed to write the story. In this study, no time limit was set and the times different participants took noted down. The study had established that the fastest writer completed the story in 17 minutes while the slowest took 24 minutes. This condition, then, required 'speeded' writing. Participants in the main study were asked if they felt pressured after completing the task and a number indicated that they did.

4. *Careful Planning (CP)*

a. Oral task

The careful planners were required to produce at least 4 sentences for each of the 6 pictures. As in the Pressured Planning condition, they were required to carry out the task after seeing the pictures for only 0.5 minutes, but they were given unlimited time to enable them to formulate and monitor their speech plans as they performed the task. Thus, the participants in this condition had limited time for strategic planning but ample time for within-task planning.

Table 4. Summary of task conditions

	Pressured planning	Careful planning
Oral Task	0.5 minutes for strategic planning; 5 minutes to tell story and required to produce at least 4 sentences for each picture	0.5 minutes for strategic planning; unlimited time for on-line planning
Writing Task	0.5 minutes for strategic planning; 17 minutes to write story and required to write at least 200 words.	0.5 minutes for strategic planning; unlimited time for on-line planning

b. Writing task

The participants were given a piece of paper and told to write the story down. They were told they could take as long as they liked and a researcher ensured that they began writing immediately. The researcher noted the time the participants spent on task to check that this was indeed longer than the time taken by the CP group. Unlike the CP group, the participants were not required to write a minimum of 200 words, as this may have been interpreted as requiring them to write quickly.

The task conditions are summarized in Table 4. It is important to note that in operationalizing the Pressured Planning condition the aim was to create a felt need to perform rapidly in the participants while at the same time ensuring that the time allocated would afford sufficient data for analysis.

Questionnaires and interviews

All the participants were asked to fill out a questionnaire in Chinese immediately after completing the task. The questionnaire consisted of open-ended questions relating to how the participants felt about the tasks and how they made use of the planning time (i.e. whether they attended to the organization of the narrative events, content or form). In addition, four participants were randomly selected from each group for a retrospective, in-depth interview in Chinese with the researcher. The participants' written responses to the questionnaire served as a basis for the questions asked in the interview. The data collected from the questionnaire and interview were used to help interpret the findings of the statistical analysis.

Measures

Measures of accuracy, fluency, and complexity were developed to evaluate the quality of the participants' oral and written production. These measures have all been used in previous studies (e.g. Foster & Skehan 1996; Wendel 1997; Yuan & Ellis 2003) and were the same for both modalities.

Planning – independent variable

1. Length of time: the total number of minutes on task was counted for each participant.
2. Syllables: the total number of syllables produced by each participant.

The purpose of the planning variables was to provide evidence that the tasks were performed in accordance with the conditions stipulated. If the length of time spent on task and productivity varied with the within-task planning conditions, this would demonstrate that the participants had performed the tasks in accordance with the instructions for each condition.

Dependent variables

Fluency Measures:

1. Production rate: the total number of syllables produced divided by the total number of minutes a participant took to complete the task.
2. Disfluencies: the total number of words a participant reformulated divided by the total number of words produced, expressed as a percentage.

Complexity measures:

1. Syntactic complexity: the ratio of clauses to T-units in the participants' production. T-units rather than c-units were used because the task performance was monologic and contained few elided utterances (see Foster, Tonkyn & Wigglesworth 2000 for a discussion of the relative merits of using T-units or c-units).
2. Syntactic variety: the total number of different grammatical verb forms used in the task. Grammatical verb forms included tense (e.g., simple past, past continuous), modality (e.g. *should*, *have to*), and voice (e.g. passive voice in the past).
3. Mean Segmental Type/Token Ratio (MSTTR). The participants' narratives were divided into segments of 40 words and the type token ratio of each segment calculated by dividing the total number of different words by the total number of words in the segment. The MSTTR (Malvern & Richards

2002) was computed for each participant by adding the mean scores for his/her segments and dividing the total by the total number of segments in his/her narrative. This procedure was followed to take account of the effect of text length on the type-token ratio.

Accuracy Measures:

1. Error-free clauses: the proportion of clauses that did not contain any error. All errors relating to syntax, morphology and lexical choice were considered. Lexical errors were defined as errors in lexical form or collocation; e.g. *I was waiting you*.
2. Correct verb forms: the proportion of accurately used verbs in terms of tense, aspect, modality, and subject-verb agreement.

The reliability of these measures was determined by a second researcher coding the data for five of the learners in each group. Pearson Product Moment correlation coefficients for the scores of the two coders ranged from a high of .99 for MTTR to a low of .82 for disfluencies, with only two below .90.

Data analysis

The normal distribution of the three groups' scores on all variables was tested by examining skewness and kurtosis. A series of one-way ANOVAs were subsequently performed followed by post-hoc Scheffe tests where appropriate (i.e. if the F score was statistically significant). Where normal distribution was not evident a Kruksal-Wallis Test was run followed by independent t-tests to compare pairs of groups. The alpha for achieving statistical significance was set at .05. t-tests were also computed to compare the two pressured and two careful groups' scores on each variable for both oral and written narratives. The Bonferroni correction was applied to safeguard against Type 1 errors. In addition effect sizes (*d*) were calculated using the formula provided by Cohen (1988; cited in Norris & Ortega 2001; 442–3) to examine the size of the effect of the different kinds of planning on performance of the task. Following Cohen, effect sizes larger than .8 will be considered 'large', sizes between .5 and .8 'medium', between .2 and .5 'small' and less than .2 negligible.

Table 5. Descriptive statistics for the independent variables (Means)

Variable	Speaking		Writing	
	Pressured	Careful	Pressured	Careful
Length of time (minutes)	3.11	4.06	17.00	21.00
Syllables	194.36	235.50	213.14	242.64

Results

To establish whether the two experimental conditions (pressured and careful production) had been successfully operationalized for both speaking and writing, the length of time spent in completing the tasks and the number of syllables produced were calculated. As Table 5 shows the tasks took longer to complete in the careful than in the pressured condition and the total number of syllables produced was also greater in this condition. These differences suggest that the way in which the participants performed the task reflected the instructions they were given as summarised in Table 4 above.

Fluency

Table 6 gives the descriptive statistics for the two fluency measures. In both speaking and writing, the participants produced more syllables per minute in the pressured condition than in the careful condition. For speaking the effect size was small ($d = .35$) whereas for writing it was negligible ($d = .01$). A different pattern emerges for disfluencies. In speaking the participants produced a greater percentage of disfluencies in the careful condition with a small effect size ($d = .36$) whereas in writing they were more disfluent in the pressured condition with a medium effect size ($d = .78$). The standard deviations for disfluencies were high, especially in careful speech, indicating substantial within group variation. Neither of the group differences for syllables per minute ($F = .88$; $p = .35$) nor for disfluencies ($F = .00$; $p = .99$) was statistically significant.

Table 6 also indicates clear differences between modalities. Thus, the participants produced more syllables per minute and were more disfluent in speech than in writing. The differences between the modalities for both variables were significant; $F = 253.49$ ($p = .001$) for syllables per minute and $F = 6.80$ ($p = .012$) for disfluencies. Independent t -tests indicated that writing resulted in significantly more syllables per minute in the pressured condition ($t = 11.93$; $p = .001$), with a large effect size ($d = 5.86$) and in the careful con-

Table 6. Descriptive statistics for oral and written fluency (Mean/SD/effect size)

Variable	Speaking		<i>d</i>	Writing		
	Pressured	Careful		Pressured	Careful	<i>d</i>
Syllables per minute	67.6 (17.01)	61.74 (17.41)	.35	12.54 (2.00)	11.81 (2.66)	.01
<i>d</i> (speaking/writing)	5.86	4.97				
Disfluencies (%)	10.51 (4.8)	13.87 (14.01)	.36	8.24 (5.73)	4.86 (2.89)	.78
<i>d</i> (speaking/writing)	.43	1.06				

dition $t = 10.60$; $p = .001$, again with a large effect size ($d = 4.97$). However, the t -tests for disfluencies failed to achieve significance in the pressured condition ($t = 1.14$, $p = .266$) and the effect size was small ($d = .43$) but approached significance in the careful condition after the Bonferroni correction had been applied ($t = 2.35$; $p = .027$) with a large effect size evident ($d = 1.06$).

To sum up, whereas modality had a clear effect on fluency, (especially syllables per minute) group condition (i.e. whether production was pressured or careful) did not have a significant effect. The interaction between modality and group condition was also not statistically significant.

Complexity

Somewhat different results were obtained for the complexity variables (see Table 7). In the case of both syntactic complexity and syntactic variety, the careful groups produced more complex language than the pressured groups. In the case of syntactic complexity the difference was statistically significant ($F = 13.94$; $p = .001$) with a large effect size evident for speaking ($d = 1.33$) and a medium effect size for writing ($d = .56$). The difference was also significant for syntactic variety ($F = 66.97$; $p = .001$) with medium effect sizes evident for both speaking and writing. However, there was no difference in lexical variety according to group condition ($F = .57$; $p = .454$) with effect sizes close to zero.

In the case of modality, the participants' written production was more complex ($F = 8.63$; $p = .005$) and more varied ($F = 6.90$; $p = .011$) than their oral production. Independent t -tests indicated a significant difference between speaking and writing under the pressured condition for syntactic complexity ($t = 4.44$; $p = .001$) but not under the careful condition ($t = 4.05$; $p = .055$). In the case of syntactic variety, the difference between both the two pressured groups ($t = 5.37$; $p = .001$) and between the two careful groups ($t = 6.27$; $p = .001$) was statistically significant. Effect sizes for both the pressured and careful groups were all large for these variables with the exception of that for the

Table 7. Descriptive statistics for oral and written complexity

Variable	Speaking		<i>d</i>	Writing		
	Pressured	Careful		Pressured	Careful	<i>d</i>
Syntactic complexity	1.33 (.14)	1.61 (.28)	1.33	1.68 (0.26)	1.92 (0.52)	.56
<i>d</i> (speaking/writing)	1.75	.78				
Syntactic variety	8.71 (3.25)	11.00 (3.42)	.69	16.21 (4.01)	18.86 (3.21)	.73
<i>d</i> (speaking/writing)	2.07	2.37				
Lexical variety	0.63 (0.07)	0.61 (.07)	0.0	0.88 (.03)	0.87 (.03)	.05
<i>d</i> (speaking/writing)	4.8	4.5				

careful groups' syntactic complexity, which was of medium size. The written narratives were also much more lexically varied than the oral narratives, a difference that was statistically significant ($F = 309.01$; $p = .001$). This proved to be the case in both group conditions. That is, in both the pressured groups ($t = 12.29$; $p = .001$) and the careful groups ($t = 12.58$; $p = .001$), writing resulted in considerably greater lexical variety than speaking. The effect sizes for modality were notably large in the case of lexical variety (4.8 for the pressured groups and 4.5 for the careful groups).

To sum up, both modality and group condition affected complexity at the level of syntax but only modality had any effect on lexical variety. No interaction between modality and group was observed for any of the complexity variables.

Accuracy

Differences in accuracy according to both group condition and modality were evident in both accuracy variables. As Table 8 shows, the careful group produced more correct clauses than the pressured group, a difference that was statistically significant ($F = 14.94$; $p = .001$). A similar result was obtained for correct verbs ($F = 13.91$; $p = .001$). The effect sizes for both these com-

Table 8. Descriptive statistics for oral and written accuracy

Variable	Speaking		<i>d</i>	Writing		
	Pressured	Careful		Pressured	Careful	<i>d</i>
Correct clauses	0.45 (0.18)	0.63 (0.13)	1.75	.77 (0.01)	.86 (0.07)	2.25
<i>d</i> (speaking/writing)	3.2	2.3				
Correct verbs	0.49 (0.18)	0.64 (0.09)	1.67	.85 (0.01)	.92 (0.06)	1.75
<i>d</i> (speaking/writing)	3.6	3.5				

parisons were large. Also, the participants' language was more accurate when writing than when speaking ($F = 60.95$, $p = .001$ for correct clauses and $F = 111.32$, $p = .001$ for correct verbs). A comparison of the two pressured groups indicated that writing resulted in more accurate language whether this was measured in terms of correct clauses ($t = 5.45$; $p = .001$; $d = 3.2$) or correct verbs ($t = 6.74$; $p = .001$; $d = 3.6$). However, the difference between the two careful groups did not achieve statistical significance, after the Bonferroni correction had been applied, for either correct clauses ($t = 7.30$; $p = .012$) or for correct verbs ($t = 5.87$; $p = .023$), although in both cases the effect sizes were still large (i.e. 2.3 and 3.5 respectively).

There was no interaction between group condition and modality ($F = 1.87$; $p = .178$). Thus, these results indicate that the participants were more accurate in their use of English when given the opportunity to perform the task carefully than when pressured to perform it quickly and, they were also more accurate in writing than in speech, especially when accuracy was measured in terms of correct clauses. Furthermore the effects of these two factors were independent of each other.

Summary of main results

Table 9 summarises the main results for fluency, complexity and accuracy. This shows that both planning condition and modality affected task performance. Careful within-task planning resulted in greater syntactical complexity and accuracy than pressured within-task planning but had no statistically significant effect on fluency or lexical variety. With regard to modality, speaking proved more fluent than writing (as might be expected) but writing was characterised by greater syntactical and lexical complexity and also increased accuracy.

Table 9. Summary of main results

Aspect of production	Planning condition	Modality
Fluency	pressured = controlled	Speaking > writing
Complexity:		
– syntactical	Careful > pressured	Writing > speaking
– lexical	Pressured = careful	Writing > speaking
Accuracy	Careful > pressured	Writing > speaking

Questionnaire and interview

The participants in the groups reported that they used the 0.5 minutes they were given prior to beginning the task to understand the pictures with some reporting that they had difficulty working out the story-lines. The participants in the oral pressured planning group reported finding formulation problematic and having little time for monitoring:

I had to think about the picture . . . I uttered whatever appeared in my mind . . . if no words came up I felt very nervous.

I realized that my sentence was wrong, but I had no time to correct it. I had to think about the next sentence. Otherwise, I couldn't finish my work in time.

The participants in the oral unpressured group also experienced problems. They reported struggling with the need to attend simultaneously to meaning and to form but they also commented that they were able to adjust their messages on-line by hesitating and reformulating. In general, the unpressured planners reported focussing more on micro- than macro-linguistic features.

Discussion

The first research question asked whether L2 learners produce more fluent, complex and accurate oral and written language when they can plan carefully on-line than when they are pressured to speak rapidly. We will consider this question by addressing the effects of the within-task planning conditions on fluency, complexity and accuracy separately.

Where fluency was concerned, the opportunity to plan carefully did result in slower production (i.e. fewer syllables per minute) in both the oral and written task but this difference was not statistically significant. The group difference in disfluencies was also not statistically significant, although a medium effect size was obtained. Thus, asking the learners to speak or write quickly had only a limited effect on their fluency. For learners such as these, whose procedural knowledge of English was limited, fluency may be much more dependent on strategic planning than on variations in within-task planning conditions (Ellis & Yuan 2004; Yuan & Ellis 2003).

In contrast, the within-task planning conditions had a marked effect on the syntactical complexity of the learners' production. Their speech and writing were more syntactically complex and varied in the careful condition, with

large or medium effect sizes evident for both variables. However, the group conditions had no effect on lexical variety in either the oral or written narratives. The explanation for these findings can be found in the theoretical models of speaking and writing outlined above. Syntactical complexity, we suggest, is a reflection of micro-planning. When learners have the time to plan on-line carefully, their propositions will be more elaborate and their formulations correspondingly more complex. However, the opportunity to plan carefully has no effect on lexical variety because this constitutes the first step in the formulation process, which learners must engage in irrespective of the time available for on-line processing. In other words, the learners in the careful group may have used the additional time at their disposal to attend to the syntactical properties of their message whereas accessing the basic lemmas needed to encode the stories was a necessity for learners in both conditions. This conclusion is supported by the learners' responses to the questionnaire and interview, which indicated a general tendency to focus on micro-linguistic aspects of production during careful within-task planning.

Careful on-line planning also promoted accuracy. The careful group produced more correct clauses and verbs than the pressured group, differences that were statistically significant and reflected in large effect sizes for both modalities. There are two possible explanations for these findings. The learners in the careful group may have been able to attend to form to a greater extent during the formulation stage. In the introduction we noted that formulation is problematic for learners with limited proficiency. Allowing learners to formulate without pressure may help them to overcome their problems by accessing linguistic resources through controlled processing. Requiring learners to perform under pressure of time may force them to rely on automatic processing. Alternatively (or in addition) they may have used the time at their disposal to monitor more. As the questionnaires and interviews showed, the learners in the pressured planning groups (especially in the case of speaking) found it difficult to monitor because of the need to attend to ongoing formulation. Thus, careful within-task planning may enable learners to draw on their explicit L2 knowledge to assist both formulation and monitoring. In short, the opportunity for careful planning frees up short-term memory, allowing learners to maximise their existing proficiency.

The second research question concerned whether there were any differences in the fluency, complexity and accuracy of the learners' oral and written productions. The results show clearly that the learners were less fluent but more complex (in grammar and lexis) and more accurate in their written than in their oral productions. With the exception of disfluencies, these dif-

ferences between the learners' oral and written productions were the same for the pressured and careful conditions. That is, careful writing was less fluent, more complex and more accurate than careful speech just as pressured writing was less fluent, more complex and more accurate than pressured speech. These comparisons produced effect sizes that were either large or medium-sized. For fluency, the ranking of the four conditions was (1) pressured speech, (2) careful speech, (3) pressured writing and (4) careful writing. For the complexity and accuracy variables the rankings were reversed. In all these cases, the differences between the two modalities were greater than the differences between the two on-line planning conditions, especially in syllables per minute and lexical variety. These results are not surprising. Writing, even when pressured, allows more time for planning, formulating, executing and monitoring messages than speaking. It makes fewer demands on short-term memory. The trade-off for reduced fluency in writing is increased complexity (lexical and syntactic) and accuracy.

The one exception to this general picture was disfluencies. Here the group condition had a different effect for speaking and writing. Thus, whereas the careful group was more disfluent than the pressured group in the oral task the opposite was the case in the written task. The explanation for this may lie in whether the learners engaged in monitoring their messages prior to or subsequent to articulation and their preference, where possible, for monitoring prior to articulation. In speech, the opportunity for careful on-line planning led to increased monitoring of articulated messages. In writing, where it is easier to monitor prior to articulation, learners may have used the opportunity for careful planning to monitor their pre-articulated messages with the result that fewer disfluencies were evident in textual output. It should be noted, however, that this constitutes only a tendency, the interaction between group condition and modality failing to reach statistical significance. Apart from this tendency, there was no evidence of any interaction between group condition and modality. Thus the answer to research question three is a clear 'no'.

A comparison of the effect sizes for the on-line and modality conditions suggests that, on balance, modality had a greater effect on the learners' productions than careful on-line processing. The average effect size for modality was 2.94 while that for the on-line processing condition was 0.88. Thus the extent to which these learners produced fluent, complex and accurate language depended principally on whether the task involved speaking or writing.

Table 10. The effects of two types of planning on fluency, complexity and accuracy

Type of planning	Fluency	Complexity	Accuracy
Strategic planning	Yes	Yes (syntactical and lexical)	Sometimes
Careful on-line planning	Limited (negative)	Yes (syntactical only)	Yes

Conclusion

Whereas the studies reported in the preceding chapters focussed on the effects of strategic planning, the study reported in this article investigated the effects of within-task planning on L2 learners' narrative productions. It demonstrates that, like strategic planning, within-task planning can have a marked effect on the fluency, syntactical complexity and accuracy of learners' output. The effects, however, are somewhat different. Whereas strategic planning has been found to promote fluency and syntactical and lexical complexity but only sometimes accuracy, the opportunity for careful on-line planning has only a very limited negative effect on fluency but leads to increased syntactical complexity (but not lexical) and accuracy. Table 10 summarises the roles of strategic and careful on-line planning.

This study was motivated in part by the wish to disentangle the effects of two factors (the opportunity for careful planning and modality), which were confounded in Ellis' (1987) study. The results show that these two factors affect learners' narrative productions separately. Careful planning promotes complexity and accuracy. Oral performance is more fluent than written performance but is less complex and less accurate. Only in the case of disfluencies was there any evidence (and only weak evidence at that) of an interaction between these two factors. Thus, we conclude that both factors work largely independently influencing output.

The results of the study are explicable in terms of the models of speaking and writing outlined in the introduction. The key to understanding the results lies in how the group and modality conditions affect the key processes of conceptualisation, formulation and monitoring. When learners have limited procedural ability in the L2 (as was the case with the Chinese learners in this study) they experience problems in formulating messages. They may be able to compensate for this lack of procedural ability by monitoring their output using explicit L2 knowledge but only if their working memories are not overloaded. Thus, the opportunity to plan carefully and to write (rather than speak) provide the most favourable conditions, allowing them additional time

to conceptualise, formulate and monitor by means of controlled processes. In contrast, pressured speaking overtaxes their working memories by requiring them to formulate rapidly by means of automatic processes and thus affords little opportunity for monitoring. The essential trade-off, then, is between fluency on the one hand and complexity/accuracy on the other, as claimed by Robinson (2001b).

Interestingly, only modality affected lexical variety. This can also be explained in terms of the models of speaking and writing. It can be hypothesized that lexical variety is mostly influenced by the elaborateness of conceptualisation. In the case of speaking (whether rapid or careful) there is little time for conceptualisation as learners must necessarily engage with the process of on-line production. In the case of writing, even in the pressured condition, there is much more time for conceptualisation (this is reflected in a much slower rate of production) and this promotes more elaborate pre-verbal plans that require richer lexicalisation.

Finally, two limitations in this study need to be acknowledged. First, the design of the study (see Table 2) meant that it was not possible to compare the same group performing the oral and written tasks under both conditions. However, given that that three groups were found to be equivalent in terms of general language proficiency, this need not be considered a major problem. The second limitation is more serious, however. The tasks used to elicit oral and written narratives were not the same. It is possible, therefore, that the differences attributed to modality were in part, at least, the result of task differences. However, as explained in the method section of this chapter, every attempt was made to ensure that the tasks were equivalent in terms of complexity.

Notes

1. There was a third condition in this study. The learners were also asked to tell the story they had previously written (after their written texts had been removed). This condition, however, is more relevant to the role of strategic than on-line planning and thus will not be considered here.
2. This discussion of interlanguage variability has been limited to the role of planning time. It is not meant to suggest that planning time is the only factor determining variability in L2 performance. Preston (1996) provides a clear account of the various other sources of IL variability, highlighting the importance of linguistic context.
3. Readers will note that whereas the same participants completed the oral task different groups of participants completed the written task. This is because the research was initially designed as two separate studies, one for speaking and the other for writing, and involved

a third condition (strategic planning), not considered in the study reported in this chapter. However, as explained in the text, all three groups involved in this study were equivalent in learning background and proficiency. Thus, there are no obvious grounds for believing that the two groups who completed the writing task were not equivalent. We acknowledge, however, that ensuring that all the tasks were completed by the same participants would have resulted in a stronger design.

4. The Higher Education Bureau Examination is an oral English examination which all students wishing to take English as a major at a Chinese university must take. The examination consists of simple questions and answers, an oral composition on an unseen topic and a story retelling.

CHAPTER 7

Strategic and on-line planning

The influence of surprise information and task time on second language performance

Peter Skehan and Pauline Foster

Chinese University of Hong Kong / St Mary's University College

Introduction

Task-based research can be linked to (at least) two sets of influences. On the one hand, and more practically, an interest in using communicative activities in language teaching classrooms leads to an interest in understanding what might make different choices of such activities more or less effective for pedagogic goals. Rather, that is, than simply reflecting on perceived success and making judgements based on unsystematic and subjective experience, researchers have sought to explore which sorts of activities, and which methods of using them, lead to more desirable measured outcomes. On the other hand, there have been developments in second language acquisition theory that regard tasks as the arena for gathering data which advance our understanding of the psycholinguistics of second language processing and second language change. Such an approach may also exploit tasks as useful research devices, and the SLA-inspired questions are likely to be of a more theoretical nature, and go beyond investigations of workability and practicality.

The most influential current psycholinguistic account of using tasks emphasises the concept of a focus-on-form (Long 1988, 1991; Spada 1997; Doughty & Williams 1998). A starting assumption here is that human beings, including second language learners, do not have limitless attention, and that the constraints of the working memory system mean that to attend to one area may well mean that other areas need to operate with less attentional resources (Schmidt 2001). A second assumption is that when attentional resources are under pressure, it is more natural for second language learners to prioritise

meaning, at the expense of form. In other words, to give a learner a communicative task to transact is likely to result in that learner responding to pressure by focussing on meaning in order to get the task done, with the result that form will only be attended to if there is spare attentional capacity and/or if something happens to direct attention to form when it would not otherwise be so directed (VanPatten 1990).

The focus-on-form approach has had a major influence upon research with language learning tasks. In effect, researchers have been exploring how different types of task and different task features might have an impact on how attentional resources are used, and therefore, how attention may be directed to form more effectively. A guiding principle is that tasks will contain some degree of naturalness of language use, and that there will be inevitable focus on meaning (Skehan 1998a). But the interest has been in how tasks, nonetheless, can support the development of control of second language learning *form* without compromising the naturalness of the meanings which are expressed. Hence the explorations of how task characteristics create the conditions for the release of spare attentional resources to be directed to form, and also how some task characteristics can bring form more directly into focus, as in Tavakoli and Skehan's (see Chapter 9) investigation of the effects of task structure.

More relevant, though, for this chapter is the impact of the conditions under which tasks are done. A number of investigators have explored how a focus-on-Form can be promoted as a result of different sets of conditions operating upon tasks, including what happens before a task, what during, and what after (Skehan 1998a). And in this area, the most researched area concerns the effects of strategic planning.

At the broadest level, there is the issue of how much time is allocated for strategic planning, without any further intervention. Mehnert (1998), for example, has shown that varying the length of strategic planning time available from one to ten minutes is associated with a move from accuracy improvement (after one minute) to complexity improvement (after ten minutes) with fluency improving by any addition of time for strategic planning within this range. Other researchers have explored whether different approaches to what is done during the strategic planning period can influence performance. For example, Foster and Skehan (1999) showed how the source of strategic planning, i.e. teacher-led vs. student-led, also had an effect with teacher-led strategic planning being more effective, and also leading to less trade-off between accuracy and complexity. Researchers have also explored whether there are interactions between availability of strategic planning time and type of task. Foster and Skehan (1996) and Skehan and Foster (1997) have reported that more com-

plex tasks lead to a greater impact when strategic planning opportunities are provided, with straightforward, retrieval-based personal information exchange tasks showing the least effects, and more complex tasks requiring justification of opinions or transformation of elements and information showing the most.

Foster and Skehan (1996) also explored the effects of different conditions operating during strategic planning time itself. In this study, in addition to a no-planning control group, they had a group who were given ten minutes strategic planning time but no suggestions as to how to use this time, and a group, again given ten minutes strategic planning time, but also instructions as to how they *could* use this strategic planning time. The study showed that the strategic planning time group with instructions were no different from the strategic planning time group without instructions for fluency; were superior for complexity; but were inferior for accuracy. Foster and Skehan (1996) interpreted these results in terms of performance tradeoffs, with the group receiving instructions being assumed to have interpreted the task in a more complex manner, such that complexity scores were raised, but with attention-absorbing implications which compromised accuracy. This result is an unusual one, and has not been replicated within the literature (but see Sanguran's study reported in Chapter 4 for somewhat different results for the effects of directing learners' attention to specific aspects of language use). We will return to our earlier study below, since this condition, although only involving one of the task types (decision-making, and not personal information exchange or narrative) from the earlier study, will be re-explored in the present study.

Wigglesworth (1997, 2001) has shown that strategic planning may have different effects at different proficiency levels, with the results in her study being more evident for more advanced learners. In contrast, Tavakoli and Skehan (Chapter 9) show strong effects at both proficiency levels in their study, with the interesting effect that the lower proficiency group with strategic planning time scored more highly on some performance measures than higher proficiency non-planners. Kawauchi in Chapter 5 also reports similar results where proficiency is concerned. These studies suggest that strategic planning can compensate for proficiency level to some degree, and help learners to use more advanced language. More generally, though, it is clear that there is scope for further research into the relationship of strategic planning and proficiency level.

A major aspect of planning which has emerged in the last few years is the distinction between strategic planning and on-line planning. Most of the earlier studies, such as those referenced here, interpret planning to involve the provision of time for preparation before a task is done. Then it is likely that the

research design of a study will explore the consequences for task performance of the pre-task phase. But more recently several researchers have distinguished between this aspect of planning and what may occur during the actual task completion phase. Wendel (1997), Ortega (1999) and Yuan and Ellis (2003) discuss the construct of on-line planning in this regard (see also Chapter 6). In other words, if strategic planning is a prospective influence upon performance, these researchers propose that in addition to the pre-task phase (which has to be, by definition, prospective), there is also the possibility that while performance is under way, second language learners may engage in a form of planning-as-regrouping, as they exploit time to enable them to “think ahead” while in the process of actually communicating. Interestingly the empirical work conducted by each of these investigators named immediately above is based upon a monologic narrative task, and so the complications that would ensue from using interactive tasks (where one contributor may have time pressures eased while the other speaker holds the floor) are avoided.

In the clearest statement of what is involved in on-line planning, Yuan and Ellis (2003) suggest that on-line planning draws upon careful production, and monitoring. Careful production, in turn, is linked to the functioning of a limited working memory capacity. Working memory, amongst other things, is used as the gateway to longer term memory, which contains both lexical and syntactic information (Miyake & Shah 1999). Yuan and Ellis (2003) propose that when communication is pressured, the working memory system only has time to access lexical information from long term memory, whereas when communication is less pressured, syntactic information also can be accessed. More broadly, and relating second language performance to Levelt’s (1989) model of speech production, they propose that both pressured and unpressured communication implicate Levelt’s Formulation stage, and so will highlight the accuracy of the language which is achieved. Yuan and Ellis (2003) also discuss speech monitoring. They discuss the sense within the Levelt model in which monitoring will require, pre-production, the speaker returning to the Conceptualisation stage when some problem is detected, and then the speech sequence is restarted. Separately, they also draw attention to Krashen’s (1981) proposals for monitoring, which do not implicate any reconceptualisation, but function more at the level of editing immediately before production. In either case, Yuan and Ellis (2003) propose that the effects of such on-line planning will be clearest in terms of the accuracy of performance, and indeed they propose that on-line planning is more relevant to increased accuracy than is strategic planning.

A central issue in planning research is that for both strategic planning and on-line planning, research designs are produced and then the effects of planning are inferred, principally through the experimental conditions involved. In other words, strategic planning is assumed to engage learners in some activities which then have an impact on performance. With on-line planning, it is the actual conditions under which tasks are done that enables the inference that on-line planning is taking place. But in neither case is there direct evidence of planning and the claims that are made are on the basis of inference only.

To put this another way, planning itself is an unobservable activity, and so has to be regarded as a construct. But what is crucial is that this construct should lead to predictions about real world performances. In the case of strategic planning, experimental conditions are manipulated (e.g. type of instructions given, source of planning, time available) and then interpretations are made as to what sort of mental operations participants were actually engaged in. Although the possibility of using more introspective approaches exists, such as post-task stimulated recall, these have been rarely used, or at least reported upon. (Mackey et al. 2000; Ortega 1999, and Chapter 3, are notable exceptions). So researchers are left to make inferences about the sorts of operations that are likely to have produced the effects which have been demonstrated. At least, though, with strategic planning, it can be assumed that the time spent pre-task has been devoted to some sort of preparation, especially if participants have been asked to make notes during the planning phase.

With on-line planning, the situation is slightly different. There is no time specifically allocated to preparation for task performance. Instead, experimental conditions are manipulated so that less pressure on communication is involved. It is then *assumed* that learners will be engaging in psycholinguistic operations which will be concerned with ongoing planning of future speech, even while current speech is proceeding. Once again, looking at things in terms of constructs, the construct of on-line planning is inferred by assuming that some attention will be devoted, under these conditions, to planning the form of future utterances. There is no direct evidence that this is happening, or that it is *planning* that is involved. As we will see below, this constitutes a weakness, and one which could be remedied by having more direct evidence available as the basis for making inferences about the mental operations engaged in.

We turn next to the issue, in task research, of how performance itself is measured. Researchers who take a cognitive view of tasks tend to conceptualise performance in terms of complexity, accuracy, and fluency (Skehan 1998b; Robinson 2001c; Yuan & Ellis 2003). Considerable progress has been made in operationalising each of these (see Chapter 1). Typically, complexity measures

have been based on the notion of communication unit, most recently in the form of the AS (Analysis of Speech) unit (Foster et al. 2000).¹ This is then used to develop an index of subordination, per unit, as an indicator of the structural complexity of speech and it is this approach that will be used in the present study.

Regarding accuracy, a range of specific measures have been used, e.g. proportion of correctly used articles, as well as more generalised measures such as the proportion of error-free units or error-free clauses, or the number of errors per 100 words. An issue which arises in accuracy measurement, though, concerns the relationship between accuracy and length of clause. Measures such as the proportion of error-free clauses provide no indication as to whether subjects are mainly using short clauses or not. It would be misleading, in other words, if a high error-free clause score were obtained by a reliance on short, possibly lexicalised phrases, such as “I think so”. For that reason, it seems worth exploring alternative measures which do take account of clause length. To achieve this, one could calculate the proportion correct of (say) three word clauses, then the four word clauses, and so on. This would give a set of numbers which, in an ideal world, would neatly reveal a cut-off point, beyond which the subject cannot produce correct clauses at whatever criterion level is adopted. Three hypothetical sets of scores are shown in Table 1.

Learner One in this table presents a clear case, in that, if we take a threshold level of 50%, we have satisfactory performance up to clauses of six words, and then accuracy falls below the criterion set. In this case, we could assign a score of 6. Learner Two is slightly more complicated, in that satisfactory performance is shown continuously until we go beyond seven clauses, except for the blip at clause length four. Here it might be proposed that a blip can be set aside, provided that it is followed by two successive satisfactory clause lengths,

Table 1. Error-free clauses and clause length: Hypothetical examples

Clause length	Percentage accuracy scores		
	Learner one	Learner two	Learner three
2	80	80	80
3	70	80	40
4	70	40	80
5	70	70	40
6	50	70	80
7	40	60	40
8	30	40	80
9	30	40	40

a criterion that is amply met in this case. Hence Learner Two would get a score of seven. Learner Three presents a more complex case. Here high-level performance is achieved, but there are never successive levels of high performance. In this case, a conservative criterion would be to award a score of the lowest level where the criterion level is met. In Learner Three's case, this would lead to a score of just two.

There is also the issue of the criterion level itself. Somewhat arbitrarily, a 50% criterion has been set. Partly this is for pragmatic reasons. Given the general level of performance from learners of the level who are typical in task research (pre- First Certificate of English, as in the present study), this criterion level generates the greatest amount of useful variance in scores. It also reflects, perhaps, the realities of working with relatively short speech samples in typical task research, with their performance emphasis, in contrast to first language acquisition research, where criterion levels like 90% are more typical. But of course this criterion could be modified, and one might also explore scores if one applied a 60% or a 70% criterion. We will explore these three criterion levels in the present research.

Regarding fluency, a number of measures have been used in recent studies. These broadly include *breakdown fluency*, indexed by the number of pauses and their length, *repair fluency*, measured by features such as reformulation, replacement, false starts, and repetition, as well as measures related to *speed of delivery*, e.g. syllables per minute, and *organisation of delivery*, measured by length of run. (See Tavakoli and Skehan, Chapter 9, for more extensive discussion of these possible measures.)

Existing measures of breakdown fluency are pause-based, and as such tend to treat all pauses in the same way. But it can be argued that pauses at the ends of clauses are more natural, and are unavoidable, and should be distinguished clearly from pauses which occur mid-clause (Freed 2000). Native speakers characteristically pause, but such pauses are end-of-clause in nature. For that reason, it may be interesting to explore what sort of relationships mid-clause pause scores have with experimental conditions.²

It could also be argued that another way in which existing measures of pausing are incomplete is that they do not typically report measures of *filled pauses*, i.e. fillers such as "ah", or "oh" or "um". Although there have been studies indicating subtle differences in the use of such filled pauses (Clark & Fox Tree 2002), more broadly they seem to function not dissimilarly to unfilled pauses in that they reflect an inability to handle speech during uninterrupted time. Measures of fluency which do not focus on these aspects of performance, too, may be missing an important way some subjects cope with the pressure

of real time communication. For that reason, a measure of filled pauses will be explored in the present study.

Having discussed measurement issues, we turn next to two variables which have not been explored in previous task research. Many studies in the task-based literature generate fairly brief samples of speech based on relatively small time periods. Typically, a recording will only contain a five-minute performance, or alternatively, and for reasons of standardisation, only five minutes of a performance will be coded and scored. This seems a rather small time slice for performance itself (not least in relation to the ten minutes strategic planning time which is often used), and this raises the question as to how performance would be different if it extended over longer time intervals. One can wonder, for example, whether the level of performance deteriorates after five minutes or whether it maintains (or even surpasses) its earlier level. In effect, to explore this would be to investigate how sustained the impact of strategic planning is. This would be useful to understand, since it might have an impact on the lengths of time that tasks are designed to enable learners to devote high levels of attention. If performance falls off after five minutes, this might well indicate that tasks would be better designed for such short time intervals and/or we might need to consider the functioning of on-line planning for task performance beyond five minutes.

In exploring further the durability of the effects of strategic planning, and the way it can facilitate performance, it is also of interest to see what sort of factors might disrupt its influence. For example, one might provide learners who are doing a task with surprise information which is at variance with what they thought was “fixed” in a task. For example, in the *Judge* task used in Foster and Skehan (1996), one of the crimes for which learners were required to agree, an appropriate sentence involved a doctor who had responded to the requests of a woman dying of cancer to assist her through euthanasia. The woman’s family had wanted the doctor to be charged with murder. This original problem could be complicated by additional information becoming available. In this case, it would be that a number of other patients of this doctor had also died through a doctor-administered overdose. In this way, one could explore what effect the provision of such additional information might have on the nature of performance, and whether those who have planned can integrate this new information into performance without undue interruption.

Drawing on this literature review, a number of research hypotheses can be proposed:

Hypothesis One: That as in Foster and Skehan (1996), *different forms of strategic planning (i.e. with or without instructions) will lead to the same selective effects on performance areas.* The motivation for this hypothesis is to attempt to provide confirming evidence for unanticipated results from the earlier study.

Hypothesis Two: That *the length of time that subjects do a task will not influence the level of performance that is obtained.* The motivation here is to explore whether performance levels can be sustained beyond short periods such as five minutes. It is assumed that the level of the learners' involvement in the task will not change and that accordingly similar fluency, accuracy, and complexity levels can be maintained as before. Ancillary hypotheses here might be to make the same prediction for all strategic planning conditions equally, rather than to expect any sort of interaction.

Hypothesis Three: That *introducing surprise elements mid-task will not have an impact upon the nature of performance and that learners will be able to absorb such additional information without performance features being affected.* It is also hypothesised that *different strategic planning conditions will be affected equally and that planners, for example, will respond to the additional information no differently to the non-planners.* The alternative to such a null hypothesis would involve learners having to regroup when presented with such new and slightly discrepant information. This might lead to the (contrary) prediction that there will be more evidence of on-line planning after surprise information is presented compared to subjects who simply carry on doing a task with the same information as before. The hypotheses (null and positive) are described in this way because this is essentially an exploratory study.

In addition to these three hypotheses, it is also intended to use additional measures, discussed above, for accuracy and fluency. With accuracy, the intention is to use a measure which relates accuracy to clause length, as indicated earlier. With fluency, mid-clause pauses and filled pauses will be recorded separately from end-of-clause pauses. This leads to another hypothesis:

Hypothesis Four: That *the new measures of accuracy will also function as good operationalisations of the accuracy construct, and that the new measures of fluency (mid-clause pauses and filled pauses) will be more effective at distinguishing between experimental conditions.* In addition, it can be proposed that the fluency data will be consistent with the existence of an on-line planning factor, a factor which should be clearly evident when on-line planning is hypothesised to be occurring.

Method

The sixty-one³ students taking part in the study were aged between 18 and 30, and attending one of six intermediate level EFL classes at an adult college in London. They came from a very wide variety of language backgrounds, and were predominantly female. All had been assigned to the intermediate level after a placement test and interview, and the particular group they attended was determined by class time (of day) and campus location, rather than any difference in English proficiency. Recordings were made of students during their normal class time as they worked in pairs on the same task. The task was to agree upon an appropriate sentence for a list of people found guilty of a variety of crimes that had resulted in the victim dying or being seriously injured. For example, a woman had killed her husband on finding him in bed with another woman. A doctor had helped an elderly woman in terrible pain to die. Another woman had tried to murder her next door neighbour in cold blood. The complete list of crimes is shown in Appendix 1.

Each class did the task under different strategic planning conditions. Classes A and B were given no strategic planning time, and had to begin as soon as they had the descriptions of the offenders and their crimes. These students would thus have to rely on on-line planning as they transacted the task. The students in the other four groups were given time before the task started in which they could engage in some strategic planning for the task ahead. Classes C and D were allowed to consider the details of the crimes, in silence, for 10 minutes before the task began, and to make notes about what they would say, though they were not allowed to use these notes during the task itself. Classes E and F were also allowed 10 minutes' silence in which to plan their ideas for the task, but these students were given some guidance notes on how they might spend this time planning what to say and how to say it. Again, they were not allowed to use their notes during the task. The instructions provided for this condition are shown in Appendix 2.

Within these strategic planning conditions was a further mid-task condition. After five minutes of task-time had elapsed, students in Classes A, C, and E were interrupted and given further details of the crimes (see Appendix 1). For example, it was revealed that the woman who had killed her husband had been in a violent and abusive marriage with a serial adulterer. The doctor was linked to the mysterious deaths of seventeen other elderly patients in the same hospital. The woman who attempted to kill her next-door neighbour was acting to avenge the death of her seven-year-old son, killed by the neighbour's dangerous driving, for which the courts had not punished him. The students had

Table 2. Strategic planning conditions for the six classes

Class A	Class B	Class C	Class D	Class E	Class F
no planning time allowed		10 minutes unguided planning time allowed		10 minutes guided planning time allowed	
extra information given after five minutes	no extra information given	extra information given after five minutes	no extra information given	extra information given after five minutes	no extra information given

to incorporate the new facts into their deliberations. These extra details were designed to derail any strategic planning the students had done at the pre-task stage, and to force them to revert to on-line planning as they took account of the new light on the crimes.

The structure of the study is set out in Table 2.

The recordings were transcribed and analyses conducted to assess the accuracy, complexity, and fluency of the language produced by the students in the different planning condition (a between-subjects condition) and also to detect any differences between the first five minutes and the second five minutes (a within-subjects comparison). Complexity was calculated by dividing the data into syntactic clauses and AS-units (Foster et al. 2000) and expressed as the ratio of clauses to AS-units. (That is, the more average clauses per unit, the greater the average complexity.) Accuracy was expressed as the percentage of total clauses that were error-free. In addition accuracy measures were computed for the proportion of clauses greater than four words long that were error-free. This was intended to provide a measure uninfluenced by any possibility that scores were inflated through the use of many short and possibly lexicalised phrases. Measures were also computed of the typical clause length that could be produced error-free, and scored using the 50%, 60%, and 70% criteria described in the previous section.

Fluency was measured through a number of indices. The main measures of breakdown fluency used in previous studies, number of pauses greater than one second, and total silence per five minutes, was separated into pauses made at the end of a clause and pauses made mid-clause. The total silence measures associated with each of these were also divided. In addition, a measure of filled pauses (e.g. 'ah', 'um' etc) was also computed, as was a measure of length of run, i.e. a stretch of language produced as a continuous whole, without pause or other sort of revision. Finally, measures of repair fluency, i.e. reformulation, replacement, false starts, and repetitions, were also computed.

Results

Each of the subject's performances was scored, for both the first 5 minute period and the second 5 minute period for a range of measures, including complexity, accuracy, and various forms of fluency. Each of these datasets was subjected to a principal components factor analysis, followed by Varimax rotation, to explore the inter-relationships between the measures. The ratio of subjects to variables entered into the analysis met the criterion of 5:1 suggested by Nunnally (1967). The results of these analyses are shown in Table 3. Some measures discussed above were not entered into the factor analysis which is reported here. This is simply to enable the factor structure to emerge more clearly. Factor analyses with all measures generated essentially the same result. Factors were extracted where eigenvalues were greater than unity, and the number of factors was not pre-specified.

Table 3. Factor analysis of task performance measures

Variable	Component			
	1	2	3	4
<i>a. First five minutes</i>				
% Accurate Clauses	–	.11	.96	–
Accuracy for clauses of five words or more	–	–	.96	–
Complexity (AS Unit)	.22	.85	–	–.14
End of clause pauses	–	–	–.19	.76
Mid clause pauses	.74	–	–	.45
Filled Pauses	.79	–	–	.11
Length of Run	–.52	.74	–	.14
Reformulation	.80	.24	–	–.24
False Starts	.78	.10	–	.23
<i>b. Second Five Minutes</i>				
% Accurate Clauses	–	–	.93	–
Accuracy for clauses of five words or more	–.14	–	.95	–
Complexity (AS Unit)	–	.68	–.46	–.17
End of clause pauses	.15	.13	–	.89
Mid clause pauses	.72	–	–	.42
Filled Pauses	.74	–	–	–.39
Length of Run	–.51	.57	–.10	.47
Reformulation	.54	–	–.31	–
False Starts	.86	.17	–	.12

Note that values below .10 are not reported simply in order to improve the ease of interpretation of the output.

The similarity in factor structure across these two analyses is striking, and so one can have some confidence that it reflects the performance on these two tasks in a stable manner. On each occasion, the first factor loads on mid-clause pausing, filled pauses, reformulation, false starts, and negatively on length of run. The pattern of loadings suggests learners who are concerned with revising the surface of language, and are actively monitoring what they are saying. Learners who do more of the component features of this then seem to produce language with shorter runs, indicated by the negative loading. It should be noted, in passing, that end-of-clause pauses, although showing a very minor loading in the Second Five Minutes data, is essentially independent of this factor.

This first factor appears to be a set of indicators entirely consistent with the construct of on-line planning. In other words, this implies that we have a group of measures which work together to operationalise on-line planning, and link particular features of disfluency to the ways on-line planning has been characterised psycholinguistically. In addition to this converging evidence, there is the intriguing discriminant evidence that pausing at the end of a clause does not link in with this on-line planning construct.

The second factor in each case concerns complexity, and tends to link AS-unit complexity with length of run. Subjects who produced more complex language also seemed to produce runs of speech with fewer disfluencies, as though, once a proposition had been identified which had some subordination involved, this unit was more often produced as a whole.

The third factor is a straightforward accuracy indicator, defined by two different but clearly related accuracy measures. The interesting issue is that there are no significant loadings of other scores on this factor. Finally, the fourth factor is a breakdown fluency factor with the highest (and defining) loading on end-of-clause pausing, with lower loadings from mid-clause pausing, and in the second five minutes, length of run. It appears, therefore, that the pausing that takes place at the end of clauses does have some connection with pausing mid-clause, but that there is distinctness also. It is interesting that for one dataset those learners who pause more end-of-clause then seem more able to produce longer runs. This may reflect them finding opportunities to regroup at the right time, as it were.

Table 4 shows the basic descriptive statistics for two of the conditions in the study, i.e. time (first five minutes vs. second five minutes) and extra information (i.e. with extra information or without extra information). The measures shown are a mixture of the conventional measures used in task-based research, and those new indices developed in this study. Regarding more conventional

Table 4. Basic descriptive statistics

	First five minutes	Second five minutes	
	All Subjects N = 61	Extra Infor- mation (N = 32)	No Extra Infor- mation (N = 31)
Complexity	1.38 (0.22)	1.36 (0.24)	1.28 (0.15)
% Acc. clauses	67.0 (13.0)	61.0 (16.0)	64.0 (15.0)
% Acc. Clauses > 4 words	55.7 (16.8)	50.40 (21.4)	51.10 (19.9)
<i>Mean word</i> 50%	7.54 (3.08)	6.31 (3.24)	5.94 (2.31)
<i>length</i> 60%	5.50 (3.02)	4.48 (2.52)	4.94 (2.08)
<i>accuracy</i> 70%	3.89 (2.42)	3.45 (1.88)	3.78 (1.98)
Pauses (end cl.)	4.05 (2.64)	4.72 (2.12)	4.13 (2.54)
Pauses (mid cl.)	6.58 (3.97)	6.14 (3.32)	6.13 (3.62)
Filled pauses	10.17 (9.25)	6.79 (5.19)	8.41 (7.13)
Length of run	5.50 (1.54)	5.50 (1.13)	5.39 (1.58)
False Starts	8.07 (3.86)	5.86 (3.45)	5.97 (4.55)
Reformulations	3.03 (2.29)	2.38 (2.31)	1.91 (1.63)
Repetitions	9.28 (7.15)	7.69 (6.31)	7.89 (5.91)
Replacements	0.50 (0.87)	0.41 (0.57)	0.41 (0.71)

Mean Scores are shown for each measure. Standard Deviations are shown in parentheses

measures, the results for complexity, error percentage, and repair fluency (re-formulations etc.) are similar to those reported in comparable studies. The only figures worthy of note are the standard deviations for the repair fluency measures, which are quite high in relation to the means which are shown. There seems to be considerable variation with these measures. The new accuracy measures (% accuracy for clauses longer than four words, Accuracy 50%, Accuracy 60%, and Accuracy 70%) pattern in a similar way to those for the conventional accuracy measure, although the relationship between standard deviations and means hints at the possibility that they generate greater variation between participants.

It is interesting that separating pausing into mid-clause pausing and end-of-clause pausing shows that more pauses are generally associated with mid-clause positions, suggesting that breakdown in performance for non-native speakers manifests itself at points other than clause boundaries. In addition, it is intriguing that filled pauses do pattern more like mid-clause pauses, with reduction in the second time period, rather than end-of-clause pauses, which show an increase.

Table 5 shows the results for within-subject t-tests comparing performance at Time 1, i.e. the first five minutes, with performance at Time 2, the second five minutes. Since there are a number of t-test results shown here, we will

Table 5. Within-subject t-tests: First 5 minutes vs. second 5 minutes

Variable	Mean: Time 1 (and S.D.)	Mean: Time 2 (and S.D.)	t	N	Signif.
Complexity	1.38 (.22)	1.31 (.20)	2.75	61	.01
% Acc.clauses	67.0 (12)	62.0 (15)	2.97	61	.01
Accuracy 50%	7.50 (2.77)	6.08 (2.77)	3.65	61	.001
Accuracy 60%	5.47 (2.82)	4.69 (2.28)	2.05	61	.05
Accuracy 70%	3.87 (2.16)	3.61 (1.91)	1.03	61	.31
Pauses (end cl.)	4.10 (2.49)	4.41 (2.35)	-0.97	61	ns
Pauses (mid cl.)	6.67 (3.90)	6.18 (3.41)	1.12	61	ns
Filled pauses	10.15 (8.50)	7.74 (6.28)	3.03	61	.01
Length of run	5.47 (1.54)	5.40 (1.41)	0.37	61	ns
False Starts	7.94 (3.72)	5.90 (4.00)	3.96	61	.001
Reformulations	2.97 (2.07)	2.13 (1.96)	3.00	61	.01
Repetitions	9.13 (6.68)	7.61 (6.03)	2.42	61	.05
Replacements	0.45 (0.76)	0.42 (0.64)	0.26	61	ns

use the .01 significance level for reasons of conservatism. Interestingly, there is a significant reduction in performance levels for the complexity measure, and for two of the four accuracy measures. It appears that in general, where a focus-on-form is concerned, learners are not able to sustain high levels of attention, and performance suffers. It is possible that with the Accuracy 60% and 70% measures, the lack of sensitivity to the measure is caused by a “floor” effect, in that at this level, all performances are quite low. Even so, the trend is in the same direction.

The findings are not so clear-cut when we compare fluency measures from the first five minutes with those from the second time period. The unfilled pausing measures, whether mid-clause or end-of-clause, do not reach significance, and interestingly, the pattern of the means sees an increase in end-of-clause pausing, but a decrease in mid-clause pausing. The pattern of decrease is also shown with filled pauses. It may be that less on-line planning engagement is evident in the second time period. Two of the repair fluency measures do show significant differences, and in each case, there is a reduction in the index concerned. Finally, there is no difference in length of run between the two time intervals.

Table 6 shows the one-way Analysis of Variance results for strategic planning, for the first time period and the second time period, taken separately. In each case, there are three levels for the ANOVA, contrasting the Control Group, the Undetailed Planners, and the Detailed Planners. In the first time period, there are significant F values for the complexity measure, two of the accuracy

Table 6. One-way ANOVAs for strategic planning

	First Five Minutes		Second Five Minutes	
	F	Sig.	F	Sig.
Complexity	5.10	.01	1.43	.25
% Acc.clauses	3.67	.03	1.68	.20
% Acc. Clauses > 4 words	3.29	.04	2.05	.14
Accuracy 50%	1.74	.19	2.57	.09
Accuracy 60%	2.48	.09	2.28	.11
Accuracy 70%	3.00	.06	1.25	.29
Pauses (end cl.)	6.19	.01	5.75	.01
Pauses (mid cl.)	2.41	.10	.41	.67
Filled pauses	2.01	.14	3.68	.03
Length of run	0.29	.75	1.14	.33
False Starts	.50	.61	.01	.99
Reformulations	1.35	.27	1.82	.17
Repetitions	.24	.79	.72	.49
Replacements	.25	.78	1.17	.32

measures, and for the measure of end of clause pausing. None of the other values reaches significance, although the significance levels for two other accuracy measures (Accuracy 60% and Accuracy 70%) are close. In other words, none of the repair fluency measures is close to significance, and the mid-clause and filled pause measures, while slightly interesting, certainly do not reach the .05 level. As it happens, it is the most regularly used measures from previous research studies which do attain significance, with the addition of the index of accuracy percentage with clauses of longer than four words. At Time 2, the only measure to retain significance is that of end-of-clause pauses. The other three measures, which did achieve significance at Time One, no longer reach even the .05 level. Slightly surprisingly, perhaps, the measure of filled pauses does reach significance at this later time period.

If we turn to the location of the significances, shown in Table 7, below, we see for the complexity measure that the Detailed Strategic planning condition outperforms the other two conditions. With percentage accuracy, the Detailed Planners (71%) score significantly higher than the No Planning condition (61%), but the Undetailed Planners are not different from either. Exactly the same relationship applies with the measure of the percentage of accuracy for clauses longer than four words. The end-of-clause measure stands out because it is the only one to attain significance at both time intervals, and in addition, the nature of the significance is identical: the No Planning condition is associated with significantly more pauses than either of the two strategic

Table 7. Location of significant results for strategic planning

	First five minutes	Second five minutes
Complexity	Detailed Planning (1.50) > Undetailed Planning (1.34) and No Planning (1.31)	ns
% Acc.clauses	No Planning (61.0) < Detailed Planning (71.0)	ns
% Acc. Clauses > 4 words	Detailed Planning (60.8) > No Planning (48.6)	ns
Pauses (end cl.)	No Planning (5.55) > Undetailed Planning (3.37) and Detailed Planning (3.35)	No Planning (5.75) > Undetailed Planning (3.59) and Detailed Planning (3.95)
Filled pauses	ns	Detailed Planning (10.74) > Undetailed Planning (6.55) and No Planning (6.00)

planning conditions, which do not differ from one another. Finally, and at Time 2 only, the detailed planners use more filled pauses than either of the other conditions.

Reflecting on the results, we can now return to the four hypotheses, which were proposed earlier. Hypothesis One predicted that the results obtained would replicate those obtained in Foster and Skehan (1996). The evidence on this point suggests that the hypothesis is not confirmed. Broadly, there is an effect for strategic planning, across all areas of performance, which is generally consistent with the previous general literature. However, the specific finding from Foster and Skehan (1996), that it would be *Undetailed* Strategic planning which would generate the highest accuracy levels, is not confirmed. Strategic planning is shown to have an effect upon accuracy, but if anything, it is the *Detailed* condition which produces the highest accuracy levels.

The second hypothesis was that *time* would not have a strong effect upon performance. This was an exploratory hypothesis and the results obtained suggest clearly that it was not upheld. Time does seem to have a marked effect upon performance, and it is clear that learners cannot sustain high levels of performance for long periods. In the present research, there is a significant reduction in performance levels after only five minutes.

The third hypothesis was that the provision of surprise information would not impact upon performance. This too was an exploratory hypothesis and was upheld, in that no clear influence of surprise information was detectable.

In fact, the confirmation of a null hypothesis here is slightly disappointing in that one might have expected differential impact of the introduction of surprise information in relation to the different strategic planning conditions.

Finally, the fourth hypothesis was simply that the new measures of accuracy and fluency would make useful contributions to illuminating the nature of the underlying constructs. There is limited support that the new accuracy measures provide some convergent validation for existing accuracy measures, and although they clearly function effectively, it is too early to say what additional information they provide statistically, even if they do make a useful conceptual contribution. The measures of fluency, distinguishing between mid-clause and end-of-clause pausing do however clearly provide additional information about the nature of performance, and their connection with the measure of filled pauses and the construct of on-line planning represents a useful contribution to our understanding of second language performance. We will explore all of these issues further below.

Discussion

In some ways, the most striking findings in this study are those which concern time, and the subjects' (in)ability to sustain high level performance. The effect is broadly similar whatever the strategic planning condition. That the effect is non-selective across strategic planning conditions suggests that having had the opportunity for strategic planning time provides no inoculation against deterioration in these areas. Interestingly, a number of the cluster of measures probing on-line planning (i.e. those which emerged in the factor analysis), show a decline in level, suggesting that the subjects' capacity to engage in this behaviour reduces, as though the capacity to reconceptualise on-line and monitor is reduced the longer the task goes on. The exception here is the mid-clause pausing measure, which does decline, but not significantly. There are no differences with the measure of automatised language performance (length of run) or of end-of-clause pauses, although there is a slight tendency in the second five minutes for *longer* pauses to come into play. Tasks, it would appear, can be demanding for learners and their capacity to sustain high-level performance is limited. Although the scale of the effect is not that great, its across-the-board nature does suggest that doing tasks takes learners to the limit, and that attentional availability for ongoing conceptualisation and formulation is finite. The implications for pedagogy are clear, in that realism of expectations on the part of the teacher has to be important.

The effects for strategic planning present an intriguing picture. First of all, a greater number of significant results are found for the effects of strategic planning during the first five-minute time-slice than the second five minutes. It seems to be the case that strategic planning conditions do not maintain their effects for long. In the first five minutes, there is evidence of the fairly standard effect of strategic planning on complexity, although the detailed strategic planning condition seems to be the major influence here, in contrast to Foster and Skehan (1996), where the undetailed strategic planning condition also led to significantly higher complexity. There is also a fluency effect, especially with respect to breakdown fluency, and the number of end-of-clause pauses that learners use. Interestingly, the putative measures of on-line planning, as defined by the first factor in the factor analysis, do *not* show any significant differences, perhaps supporting the claim of a separation between the two types of planning.

The key area here, though, concerns accuracy, since Foster and Skehan (1996) report an accuracy effect, but with *undetailed* planners producing more accurate language than did the detailed planners. In the present study, there is also a clear accuracy effect, but it is slightly different. Looking at the results for Time One in Table 6, there are significant results for two out of the five accuracy measures (i.e. percentage of error free clauses, and percentage of error free clauses longer than four words), and three measures approach but do not attain significance (Accuracy 50% (.19); Accuracy 60% (.09); and Accuracy 70% (.06)). None of the values at Time Two attains significance. Importantly, however, when one looks for the location of this effect through post-hoc tests for both the significant measures at Time One, the contrast is between the two strategic planning conditions, on the one hand, and the no planning condition, on the other, with the no planners producing significantly less accuracy. Interestingly, for both percentage error-free clauses and error-free longer clauses, the highest accuracy score is found with the *Detailed* strategic planning condition, in contrast to Foster and Skehan (1996), where it was the undetailed condition that generated most accuracy.

The results for the introduction of surprise information mid-task are a little disappointing. The effects of this condition are not terribly strong. Regarding accuracy, the effect of additional information does not distinguish the performances in the second five minutes between the non-extra information condition and those who had the extra input on the judgements they had to make. There might be a slight tendency towards extra information leading to *less* reduction in accuracy in some of the measures, such as Accuracy 50%, but this was not a strong or even consistent trend. With fluency there were tenden-

cies for the end-of-clause and mid-clause measures to be higher in the second five minutes for the extra information group, suggesting that they did need to regroup more. This tendency was not apparent with the filled pauses measure, and so it is difficult to judge whether the extra information group were induced to engage in more on-line planning. Finally, with complexity, there were no clear trends. Except for the non-planners, who were not disadvantaged by the extra time period, the two strategic planning conditions, with or without extra information, saw a reduction in performance in the second five minute period.

We can now reflect on these results more broadly, and relate them to the literature. What we have learned that is distinct from previous studies is that performance on tasks is sustained at a high level with some difficulty. Even within what might be considered quite brief time periods, the detailed performance measures show reductions. This is interesting, and has implications for research and pedagogy. Regarding research, it suggests that the length of time during which tasks are done will need to be considered more widely, however the task is conceptualised, and however it is measured. Regarding pedagogy, if the purpose of giving tasks to learners is to push them to useful performances which develop their interlanguage or their abilities to use language, it would appear that attention cannot be sustained easily for extended periods. This may suggest a need for realism when tasks are used. It may also suggest that post-task activities may be valuable to consolidate the fleeting effects of the manipulations which are used.

The present study has also made some contributions by introducing new measures of task-based performance. The new accuracy measures have turned out to make a slight conceptual advance, in that they avoid a limiting dependence on all clauses, treated equally, as the unit of measurement. Instead, they relate clause accuracy to clause length, and in this way, provide what might be termed a power index. This increases not when clauses-in-general are handled without error, but only when clauses of greater length are more correctly produced. It is perhaps worth noting, in this regard, that in the factor analyses, all the accuracy measures loaded together, and were distinct from the complexity measure. In other words, it does not appear to be the case that as accuracy is linked to clause length that it becomes a disguised complexity index. With this data, the two aspects of performance are distinct. In the event, there are no obvious additional statistical contributions made by this range of power measures, but it is clear that they do merit further scrutiny.

More relevant, perhaps, is the separation of pauses into those that occur end-of-clause and those which occur mid-clause, as well as the counting of filled pauses. The factor analysis shows that the mid-clause and filled pause

measures load on the first factor, as do reformulations and false starts. Further research is clearly needed, but it could be here that we have the beginning of a measure of on-line planning, in that all these measures, in addition to cohering statistically, reflect a more active involvement with the surface of language while it is being produced. Such a measure may actually enable on-line planning to be operationalised more effectively.

But even this is not without its problems, since on-line planning may turn out to be very difficult to delineate precisely. It would seem, for example, that while speech is actually being produced, attention could be directed to (at least) four areas:

- a) to snatching time during on-line performance in order to plan the language for what will be said in a clause or two. Such an allocation of attention would be prospective (as in strategic planning) but would occur during performance,
- b) to buying time to reconceptualise the content of what will be said. Again this would be a prospective use of attention.
- c) to monitoring ongoing performance so that something which might have been said (possibly erroneously) is not,
- d) to reacting to a communication problem whose existence has just been realised.

The first of these relates most clearly to Yuan and Ellis' (2003) careful production. Their formulation, though, implies that this approach to speech production is a response to experimental conditions, such that time is available, and used, not only to access lexical information in long-term memory, but syntactic information also, and assemble this information in a working memory buffer. In the present case, it is, slightly contrastingly, being hypothesised that, although some conditions may make careful production more likely, it is possible for any second language speaker to try to protect time and think ahead, whether linguistically or conceptually, as in a) and b) above, with b) perhaps corresponding to Levelt's restarting of the Conceptualisation-Formulation-Execution sequence. The third possibility above corresponds to post-production monitoring (Krashen 1981; Yuan & Ellis 2003) whereas the fourth concerns the speaker who simply runs up against an unforeseen (and possibly major) difficulty.

The problem we have in working with on-line planning is that, on the one hand, we have a range of putative psycholinguistic processes which implicate on-line attention to speech (not all of which may be planning), and on the other, we have a range of performance measures. Currently on-line planning

studies draw on experimental conditions to justify the claim that on-line planning has been operative. Ideally, however, we would have a theory to link the putative processes with the different measures available. It is to be hoped that the data reported here, if not remotely connecting with such a theory at this stage, will at least provide important input to such a theory. We certainly do have a range of measures which cohere (mid-clause planning, filled pauses, reformulation etc.) What we need next is to be able to relate these (and possibly other measures) to the construct of on-line planning.

Summary and conclusions

The present study partially replicated Foster and Skehan (1996), while adding additional task conditions, in order to try to understand better the way attention is used during tasks, and how different performance areas may be differentially affected. The results of Foster and Skehan (1996) were partially replicated, although with some discrepant results. In addition, the effects of giving learners more time to work on a task, beyond five minutes, were clearly more salient than those of giving task completers some additional discrepant information. The study therefore adds to the data on the effects of strategic planning, and also shows how the time period during which a task is done may have a strong influence on the results.

One conclusion to draw from this is that we need further research to understand the role of strategic planning. Learners are clearly doing different things during this planning time, whether that is the result of personal idiosyncrasy, or that of manipulated experimental conditions. We currently do not know what is going on in this period, and it may be that more qualitative approaches will need to be used (see Ortega, Chapter 3), in a more exploratory manner, to enable progress to be made.

But the other general conclusion is that there is scope to develop more direct measures of on-line planning. At the very least, the use of additional measures will sharpen up our understanding of the construct. At present we have to take on-line planning to be a measure of how much speakers regroup in real-time as they modify what is formulated as their utterance. This may be a useful perspective, but it needs to be supplemented with more direct manifestations of the psycholinguistic process that must be involved here. The discussion of measures of on-line speech compensations, such as filled pauses and mid-clause (rather than end-of-clause) pauses, provides some suggestions as to how

progress can be made in this area. Without such research, our understanding of the construct of on-line planning will be limited.

Notes

1. These are similar to T-units and c-units but with a much more detailed definition that allows 'messy' spoken data to be coded in a principled way. They divide spoken language into syntactically complete units and can be used to calculate how frequently certain features occur in speech samples. Extensive discussion and examples are provided in Foster et al. (2000). See also Chapter 9, this volume.
2. Ellis (personal communication) makes the excellent point that there may also be differences between native and non-native speakers for end-of-clause pauses. Native speakers may indeed pause here, but non-natives may pause longer. To address this issue, native-speaker baseline data would be required, and this is not available for the present dataset.
3. Originally there were sixty-two participants, working as thirty-one dyads, but one voice proved very difficult to transcribe accurately and was accordingly omitted from the analyses.

Appendix 1

The instructions were given to the Surprise Condition after five minutes. These include the original crimes.

Before you make your final decision, here is some extra evidence on each of the cases. Do you want to change your judgement because of this new evidence? Why? Why not?

1. The accused found her husband in bed with another woman. She took the breadknife and killed him.

The husband was a violent and abusive man. He has often attacked his wife and children. He had many girlfriends, and this was not the first time his wife had found him in bed with one of them.

2. The accused is a doctor. He gave an overdose (a very high quantity of a painkilling drug) to an 85-year old woman who was dying painfully of cancer. The doctor says that the woman had asked for the overdose. The woman's family accuse him of murder.

Seven other people in the same hospital have died in a similar way, through overdoses. The doctor refuses to say if he was involved.

3. The woman is accused of trying to kill her neighbour whom she hated. She bought a gun, followed her neighbour to a park and shot him. He was badly injured, but did not die.

The woman hated her neighbour because he was driving the car that had killed her 7-year old son. The neighbour had only one eye, and no driving licence. He was charged with dangerous driving and sent to prison for only three months.

4. The accused is a professional footballer. During a match he kicked another player and was sent off. When he was leaving the field, he attacked a spectator, kicking him in the face and breaking his teeth.

The footballer is black. The spectator is a member of a Neo-Nazi organisation. Other people at the match say that the spectator was with a large group of people who were shouting racial abuse at the footballer. This group goes to every match and always shouts racial abuse at this player.

Appendix 2

Detailed planning instructions for the “Judge” task

You have ten minutes to prepare for this task. You can make notes during the ten minutes, but you won’t be allowed to use these notes while doing the task. Be sure you can explain the decisions that you make to your partner.

These are things you can do to help you prepare.

- think what you already know about each of the questions for judgement
- think about the different parts of each problem
- think of different judgements that are possible; of reasons for these different judgements; and of why these reasons could be wrong
- think about why your partner may not agree with you
- think what grammar you need to do the task
- think what vocabulary you need to do the task
- think how to avoid difficulties and solve problems with grammar and vocabulary
- think how you will deal with *talking* to someone, how you will say what you want to say, and stop the other person from talking all the time

Make your notes here

Name: _____

(Space provided here for note-making)

Use the back of the paper if you want to make more notes

SECTION V

Planning in language testing

Whereas the previous studies have all examined the effects of planning on task performance in either a classroom or a laboratory setting, the two studies in this section examine its effects on L2 learners' performance of tasks in a testing context.

It is worth asking what actually is meant by a 'testing context' and how this might differ from a classroom or laboratory context. The most obvious difference is that testing constitutes a 'high stakes' context, where how the learners perform the tasks 'matters'. We do not know how this influences the way in which learners orientate to tasks but it seems reasonable to suppose that (1) a testing context will lead to increased attention to accuracy with concomitant reduced attention to fluency and complexity and (2) the felt need to attend to accuracy during performance through careful on-line planning may override any effect for strategic planning on either accuracy or other aspects of performance. It is also worth noting, however, that there can be marked differences from one testing situation to another and that learners may also differ markedly in how they orientate to a testing context.

As we noted in Chapter 1, previous studies of the effects of planning on task performance in a testing situation have produced mixed results. In general, however, planning seems to have less effect on performance on tasks in tests than in other contexts. These mixed results are reflected in the two chapters in this section. Elder and Iwashita failed to find any effect for strategic planning (3 minutes) on any aspect of performance. Nor did they find that the provision of planning time had any noticeable effect on learners' perceptions of task difficulty. In contrast, Tavokoli and Skehan found that strategic planning (5 minutes) had clear effects on fluency (but only 'breakdown fluency', not 'repair fluency') and on accuracy and complexity. Like Kawauchi (see Chapter 5) they also found an interaction between planning and learner proficiency, such that in some instances it was the less proficient learners that benefited most. Tavokoli and Skehan also found that the testees given time for planning reported finding both the structured and unstructured tasks in their study easier than those learners who had no planning time.

Why are the results of these two studies so different? Clearly, there are many possible reasons. Elder and Iwashita provide an excellent account of the various possibilities in the discussion section of their chapter. It is also clear, as Tavokoli and Skehan conclude, that this is an area in need of further research.

My own guess is that the different results reflect the fact that the testing contexts in the two studies are really very different. Elder and Iwashita's testees completed a whole battery of tasks as a group in a language laboratory with the tasks used to measure the effects of planning coming 6th or 7th in the sequence. Further, these testees knew that they would be subsequently taking an institutional TOEFL test. In short, my guess is that these learners clearly perceived that they were being assessed. In contrast, Tavokoli and Skehan's testees only completed four tasks, for two of which they were given planning time. They were specifically told they could take notes during the planning time, as in other 'pedagogic' studies of tasks. They performed the tasks in a language laboratory but they did so one-on-one with a researcher. They were not asked to complete any other test. In short, the conditions of task performance in the Tavokoli and Skehan study resembled those in a standard task laboratory study much more closely than those in the Elder and Iwashita study.

We are left, then, with the question I posed above; what do we mean by a 'testing context'? Until we have a clearer answer to this it seems very likely that we will continue to find that testees respond very differently to the opportunity to plan in different testing studies.

CHAPTER 8

Planning for test performance

Does it make a difference?

Catherine Elder and Noriko Iwashita

Monash University, Melbourne / The University of Queensland, Brisbane

Introduction

The role of planning in language production has long been acknowledged in studies of both first and second language acquisition. Underpinning some of these studies are theoretical models of information processing as formulated by McLaughlin (1987) and more recently Van Patten (1990). These scholars characterize spoken performance as drawing heavily on attentional resources, with the result that one aspect of speech production takes priority over others, especially in the case of learners with limited proficiency. The provision of planning time has the potential to free up the learners' attentional resources so that they can focus more closely on features of the linguistic output thereby enhancing the quality of their performance.

A somewhat different take on the process of L2 speech production is provided by Anderson (1983, 2000) who sees learning as involving the transformation of declarative knowledge (about the language) to procedural knowledge (ability for use). Accessing and applying declarative knowledge requires conscious effort, particularly for second language learners, and is likely to slow down linguistic production. Taking time to plan before speaking may allow access to declarative knowledge through preparation and rehearsal strategies, enabling specific lexical and grammatical forms to be retrieved and assembled during task performance and/or allowing ready made utterances to be recalled and reproduced.

An alternative to Anderson's two types of knowledge model is advanced by Skehan (1998b), who distinguishes between rule-based and exemplar-based systems. The latter, comprised of chunks or formulae, is easily accessible,

whereas the former is more demanding in processing terms. The rule-based system may be activated when there are adequate amounts of processing time, allowing for more creative and accurate language use. Skehan has argued that second language teachers' choice of language learning tasks should be made in a principled fashion, with various aspects of the task manipulated to provide opportunities for different types of language use (e.g., spontaneous versus rehearsed production). He also points to the value of research into the effect of these different task conditions and task characteristics on performance for language testing:

The conditions under which tasks are done and the way conditions interact with performance are a fertile area for research. Already we can see how, in areas such as strategic (test) planning, there are findings which clarify how conditions of task elicitation influence performance. If we are to understand how testing conditions are to be standardized and/or how an adequate range of sampling conditions is to be identified, an understanding of this set of influences is essential. Unless research bearing on these factors becomes available, generalizations based on actual test performance will be a hazardous and chance-dominated undertaking. (1998: 177)

Bachman (1990, 2002), although supportive of this kind of research, takes issue with Skehan's approach to characterizing task difficulty, pointing to confusion between characteristics of the task and attributes of the test taker in Skehan's model of task complexity. According to Bachman, some aspects of complexity identified in Skehan's model (e.g., cognitive complexity and communicative stress) are not inherent in the tasks themselves, but instead depend on the affective reactions and cognitive capacity of the test-taker. Task difficulty, in his formulation, is the product of an interaction between many facets of the assessment including both features of the task and test taker abilities and affect (2002: 466). Nevertheless, there is general consensus in the field of language testing as to the importance of exploring the relationship between task characteristics and test performance and there are now numerous published studies looking at one or other facet of test tasks or items (e.g., Norris, Brown, Hudson & Bonk 2002; Brindley & Slatyer 2002) and their influence on performance. Studies which focus particularly on the effect of planning time on test performance are however few in number, the most well-known being that of Wigglesworth (1997).

In addition to the analyses of test scores, Wigglesworth (1997) examined test candidates' discourse in various types of task performance (i.e., picture description, narrative story retelling) as investigated in SLA studies (e.g. Crookes 1989; Ellis 1987; Ortega 1999). She compared the accuracy, fluency and com-

plexity of candidate test discourse under two conditions (i.e., with and without strategic planning time). The results showed that test candidates who were given 1 minute of planning time improved their performance in terms of complexity (amount of subordination), fluency (number of self-repairs) and accuracy (suppliance of plural-s, verbal morphology and indefinite article). While Wigglesworth's findings on the effect of planning time on test performance were somewhat mixed (i.e. the effect was not consistent across tasks and also varied according to the proficiency level of test takers), they nevertheless gave tentative support to Skehan's hypothesis that planning time may be associated with more complex, more accurate and (to a lesser extent) more fluent speech. However, in her study the impact of planning was visible only in some features of the candidates' discourse produced under the strategic planning condition and did not result in raters assigning higher test scores to candidate performance. She speculates that raters, who rely on impressionistic judgments guided only by the generic descriptions contained in rating scales, may be unable to make the fine distinctions yielded by discourse analytic methods.

Building on her 1997 study, Wigglesworth (2000) further investigated the effect of planning time in conjunction with another variable (i.e., task familiarity or structure¹). In total 400 adult migrants studying English in Australia at five different levels (80 learners at each level) participated in the study. Task One served as a control, with all participants taking it under the same condition. For Tasks Two to Five, 20 learners in each level took each combination (i.e., a structured task with and without five minutes' planning time and an unstructured version with and without planning, a familiar task with and without five minutes' planning time and an unfamiliar task with and without five minutes' planning time). Wigglesworth sought data from three sources: a) analyses of test-taker discourse, b) quantitative analysis of test scores (using both the raw scores assigned by raters and the adjusted scores yielded from a Rasch analysis) and c) test candidates' evaluation of task difficulty. Results showed that, contrary to expectation, candidates received a higher test score on a familiar task where no planning time was given, but that planning had no effect either way for the unfamiliar task. Interestingly, planning time appeared to adversely influence performance on both structured and unstructured tasks (although its impact was small). She speculates that there may be a conflict between what raters attend to in their marking (i.e., mainly the accuracy of production) and the focus of learner activity during planning (i.e., the content of what they are going to say) and that in actual performance the latter may be at the expense of the former. As for the evaluative feedback, the learners perceived the familiar task to be easier with planning, but planning time did not seem to have

much impact on the perceived difficulty of either the unstructured or structured task. Because planning time was manipulated with another variable, it is hard to tease out the effect of planning time *per se*. Nevertheless it appears that, although greater amounts of planning time (i.e., five minutes) were provided than in the previous study (Wigglesworth 1997), this did not produce a marked effect on performance either at the discourse or rating score level. Where differences occurred they were not in the predicted direction.

These latter findings conflict somewhat with those of Mehnert (1998) who specifically investigated the effect of different amounts of planning time (i.e., none, one minute, five minutes and ten minutes) on the speech performance of L2 (German) learners. Several aspects of learner discourse were investigated (i.e., speech rate, mean length of turn, S-nodes/T-unit, errors per 100 words, and weighted lexical density). Mehnert found that fluency of speech improved as planning time increased, but that gains in complexity were achieved only under the longer planning time condition (i.e., ten minutes). Accuracy behaved somewhat differently, with more accurate speech observed only under the 1 minute planning condition. Nevertheless, although the relationship was not always a linear one, planning time was found to have a significant effect on all three aspects of speech production. The unambiguous relationship between accuracy and planning time was also found in several other planning studies conducted in non-testing contexts (e.g., Foster & Skehan 1996; Wendel 1997) so the generally inconclusive findings of Wigglesworth's research are somewhat unexpected. The impact of planning time clearly warrants further investigation, along with other variables which may influence information processing and language production in language testing contexts.

It is worth mentioning at this point that few of the well-known public tests of oral proficiency allow for substantial amounts of planning time before candidates start speaking. Where planning time is allowed on standardized proficiency tests, it is usually for the sake of fairness, to make sure that candidates have a chance to digest information presented in a prompt (e.g., map, graph etc.) so as to be able to use this information in their subsequent speaking performance. For example, the latest version of the Test of Spoken English (TSE) allows a maximum of one minute's preparation time, but only on tasks which require learners to process graphic and verbal input in pictures and other diagrammatic stimulus materials. Likewise, in the International English Language Testing System (IELTS) test, one minute's preparation is allowed for only one task (Task 2) so that candidates will be able to absorb some written input, but in other tasks test-takers are not expected to read or listen to any information, and therefore no planning time is given. The TOEFL Academic Speaking

Test (TAST) is even more parsimonious, providing a mere 30 seconds' strategic planning time for task requiring candidates to synthesize material provided in both written and oral form. In all three tests the time allowed for these more formal lecture-type tasks is far less than a learner would have in preparing for a corresponding 'real world' oral presentation in an academic or professional context. The reasons for this may be largely practical. Given the constraints of the testing situation there is always a trade-off between the time required to elicit a rateable sample of speech on individual tasks and the number of tasks required to achieve adequate coverage of the domain of interest. Nevertheless this places obvious limits on the authenticity of such tasks as measures of academic speaking ability, where planning prior to presentation is often the norm. Moreover it would seem that decisions as to how much planning time to allow in tests of speaking ability, or indeed whether to allow any time at all, are generally made on an *ad hoc* basis with little attention to psycholinguistic models of speech production and to the potential guidance such models might offer for the design of test tasks.

This limitation was acknowledged in a review of the speaking proficiency construct conducted as part of the TOEFL 2000 project (Butler et al. 2000) and a study was commissioned with the aim of operationalizing Skehan's information-processing model of task complexity (Skehan 1998b) in a language testing environment. The project explored the effect of a number of different task dimensions, including planning time, on test performance (see Iwashita, McNamara & Elder 2001, for an account of the findings). The aspects of the investigation relating to planning time have been extracted here for further consideration.

The part of the study we report in this chapter is similar to that of Wigglesworth (1997) in that it considers the effect of strategic planning on monologic performance in the context of a tape-based test of speaking proficiency. As in Wigglesworth (2001), we also gathered data on test taker perceptions of the task conducted under planning and no planning conditions, based on the assumption that these perceptions give some insight into what test takers bring to the task. We also believed that attitudes towards a task and the conditions under which it is performed might itself have some impact on test performance, particularly when the test is tape-based and there is no flexibility in the way the test tasks are delivered. (For a brief review of studies into test taker perceptions see Elder, Iwashita & McNamara, 2002.) Indeed Robinson (2001c), who has conducted a similar study in a non-test environment, has proposed a distinction between complexity (a feature of the task) and difficulty (operationalized in terms of perceptions of task difficulty on the part of learners) and has found,

as he predicted, that both of these are related to the quality of speech production. While his conceptualization of “difficulty” is different from the way the term is defined in the measurement community, the notion that task performance involves an interaction between features of the task and characteristics of the test-taker is beyond dispute (Bachman 2002) and is the cornerstone of widely used measurement models such as Rasch (McNamara 1996).

Research questions

Three research questions were addressed in this study:

1. Does the provision of strategic planning time make a difference to oral language production as reflected in scores derived from ratings assigned to test-taker performance?
2. Does the provision of strategic planning time make a difference to language production as reflected in features of the oral discourse produced by test-takers?
3. Are differences in the amount of strategic planning time provided associated with differences in test taker attitudes and perceptions of task difficulty?

Method

Design

The speaking test was administered in a university language laboratory. All participants were randomly assigned to one of four experimental groups as summarized in Table 1. They completed a language background questionnaire, followed by the speaking test. They later took an Institutional version of the TOEFL test (so that we could control for any differences in ability from one group to the other). The speaking test was made up of eight narrative tasks (3 minutes maximum for each task) covering the four different cognitive dimensions manipulated in the larger study (i.e., ‘Immediacy’, ‘Adequacy’, ‘Perspective’ and ‘Planning’, the latter being the only one of interest here). Participants were granted a 10-minute break after the first four tasks. The order of presentation of the + and – (i.e., more or less cognitively demanding) conditions of each dimension was counterbalanced across the four experimental groups, with the “Planning” tasks administered after the break in either 6th or

Table 1. Design of the present study

Group 1 (n=37)	Group 2 (n=58)	Group 3 (n=51)	Group 4 (n=47)
Version 1+	Version 2–	Version 2+	Version 1–
Version 2–	Version 1+	Version 1–	Version 2+

Note: + with planning time, – without planning time

7th position as indicated below. Under the “+ planning time” condition, participants were given 3 minutes’ planning time in addition to 75 seconds to read the instructions and look at the pictures, whereas under the “–planning time” condition, participants were given only 75 seconds to read the instructions and look at the pictures. It was hypothesized (in line with previous research) that a story told after 3.75 minutes’ planning time would be less difficult i.e., it would be associated with a more fluent and (possibly) more complex candidate response than a story told with only 0.75 minutes’ planning time (i.e., the minimum amount we reckoned to be fair if candidates were to be able to make sense of the task). Predictions relating to accuracy were harder to make, given the mixed results of previous research (see Introduction to this volume), but Skehan’s model predicts that provision of planning time at least has the potential to encourage focus on form thereby reducing the incidence of error (Skehan 1998).

We acknowledge that the amount of planning time (3.75 minutes) given to our participants is much less than in many SLA studies on planning time (i.e., 10 minutes), but given the practical constraints on time in any operational test situation noted earlier, we felt that was as much as could be feasibly allowed, if the findings were to be applied to a tape-based testing of speaking of the kind envisaged for the Next Generation TOEFL.² (Note that practicality is a key principle underpinning all language testing practice (e.g., see Davies 1990)).

After completing each task, test-takers completed a one-page questionnaire on their perceptions of the task. Their responses to two of these questions, one relating to task difficulty, and the other relating to task enjoyment, are reported below.

Participants

197 students were recruited for this study (male = 75, female = 122) on the understanding that they would be taking part in two tests: a multi-task test of their speaking proficiency, followed by an Institutional version of the TOEFL and that they would be given feedback on their performance. The majority

(approximately 80%–90% of the participants) was currently enrolled in an ESL course in Melbourne to prepare for study at university in Australia, while the remainder was already studying at a tertiary institution in Melbourne. The mean age of the participants was 21.6 years ($SD = 4.5$), and the mean length of residence in Australia was 4.3 months. The L1 of participants varied, but the majority were speakers of Asian languages (e.g., Chinese, Vietnamese or Japanese). The mean length of time they had been studying English was 6.9 years; many had also studied foreign languages other than English at some time. Most participants spoke English at home in Australia.³ The mean score of the participants on the Institutional version of the TOEFL test was 493.1, with a SD of 45.8, and a range of 427 to 670.

Materials

1. *Tasks*

Speaking tasks used in the study involved a single type of stimulus, of the kind used routinely in the Test of Spoken English (TSE), namely: a narrative task based on a sequenced set of picture prompts. Two exemplars of narrative tasks (1 & 2) were used to investigate any effect for specificity of task or, in other words, to determine whether our operationalization of the experiment was generalizable across different task exemplars. Expert judgments (i.e., consultations with experienced teachers and researchers who have used narrative tasks in their classroom teaching and research respectively) were canvassed in an attempt to ensure that the two exemplars of each task dimension resembled each other as closely as possible in terms of their linguistic demands and likely level of familiarity to the test takers. The tasks were piloted and revised before embarking on the main study so that any unforeseen problems associated with either task could be remedied. One of the tasks used for this experiment is reproduced in Appendix A.

2. *Questionnaire*

All participants were asked to complete a questionnaire after each task. This contained questions about their perceptions of the difficulty of each task, and their attitudes towards it (defined in our study as enjoyment). Answers were given on a 5-point Likert scale. In the analysis of responses, answers were coded on a scale of 1 to 5 with 5 representing the most favourable response (easiest, most liked) and 1 the least favourable (hardest, least liked). Examples are:

Perceptions of the difficulty of the task

Q1 *Did you find the task easy or difficult?*

Attitudes towards the task

Q5 *I enjoyed telling the story*

These questions were designed to tap test-takers' reactions to the story-telling experience, without drawing their attention to the particular condition under which it was performed.

Analysis

1. *Assessors' ratings*

Performances of all subjects were rated using analytical rating scales for fluency, accuracy and complexity specifically developed for the study (see Appendix). In total, fourteen raters were recruited for the assessment of the speaking tasks. All raters had some experience in rating speaking tests (e.g., IELTS, Occupational English Test for medical professionals, TSE) as well as teaching ESL at a level similar to the participants in the present study. Before assessing the speaking tasks, all raters participated in the rater training session, and then were asked to rate sample tasks for accreditation. Each of the performances received two independent ratings from any pair of raters drawn from the pool of 14.

These data were analyzed using the IRT based program FACETS (Linacre 1990) in order to determine whether there was any impact of the imposed conditions on the scores assigned by raters to task performance. (The methods used to minimize the effects of any individual rater inconsistencies are described in Elder et al. 2002:356). Suffice it to say that the error levels associated with the difficulty estimates (known as logits) yielded by the IRT analysis are acceptably low). The impact of the imposed conditions was evaluated by means of a *t*-test for differences in the estimates of the difficulty/ability (logits) presented by the + and – Planning conditions. This analysis was conducted with adjusted difficulty estimates (derived from the fluency, accuracy and complexity ratings combined). Since it is not possible within the Facets program to run separate analyses for each of the fluency, accuracy and complexity criteria, further *t*-tests comparing + and – Planning conditions were run using the average of the two raw scores awarded to each candidate on each of these criteria respectively. These raw scores are however far less accurate measures of ability than the logit values yielded by the Facets program (which are adjusted to compensate for differences in both rater severity and the difficulty of task versions). Any difference in the difficulty of the versions is not reflected in the raw scores. The reliability estimates for each rater pair also vary widely and this

may be partly a reflection of the very small number of candidates assessed by each pair. Given these limitations, the results of this secondary analysis must be interpreted with caution.

2. *Discourse analysis*

A subset of 36 subjects was randomly selected from the larger group, but minor adjustments to the sample were made to produce equivalence across conditions in terms of gender and TOEFL score. The number of participants, while somewhat arbitrary, was chosen for the following reasons: 1) it constituted a little under 20% of the entire data; 2) the number needed to be limited because of the work involved in transcription; 3) the data set, while not large, is significantly larger than that reported in most previous studies where transcripts of between 10 and 20 students for each dimension are the norm. The data were transcribed using standard orthography, and entered using the computer program from the CHILDES project (MacWhinney 1995). The measures used in Skehan and Foster (1999) and Foster and Skehan (1996), with minor modifications, were adopted in analysing the test discourse. These measures are:

Fluency, operationalized as the number of repetitions, false starts, reformulations, and hesitations and pauses, divided by the total amount of speech (measured in seconds)

Accuracy, operationalized as the percentage of error-free clauses in the total number of clauses

Complexity, operationalized as the number of clauses divided by the number of c-units⁴

Inter-rater reliability. Inter-rater reliability for coding was calculated by getting an independent researcher to code 10% of the data resulting in 91% level of agreement.

3. *Test-taker feedback (Questionnaire data)*

First of all, mean scores were calculated for each question under the two conditions in each task. *t*-tests were then carried out to examine whether the provision of planning time made any difference to candidates' perception of task difficulty and their level of enjoyment as reported immediately after telling the story.

Results

Findings relating to each of the three research questions are summarised in Table 2 below, together with a description of the types of data used and the analyses undertaken. A more detailed report on these results is offered below.

RQ 1 Does the provision of strategic planning time make a difference to oral language production as reflected in scores assigned to test-taker performance?

This question was answered by means of a quantitative analysis of test scores derived from assessors' ratings. The results, based on comparison of adjusted ability estimates (logits) in the + and – condition, were non significant ($t = 0.67$), indicating that the provision of planning time made no difference to the scores achieved by candidates (see Table 3), or, in other words, that provision of planning time made no difference to task difficulty when fluency, accuracy and complexity ratings were combined to produce an aggregate score of the kind that is currently used as a basis for reporting performances on the TSE and most other tests of English for academic or professional purposes. The additional *t*-test analyses which were performed using the raw scores awarded for fluency, accuracy and complexity respectively also yielded non significant results (see Table 4) indicating that there was no effect for planning time on any of these different aspects of performance.

Table 2. A summary of methodology and results

	Assessors' ratings	Discourse analysis	Test-taker feedback
Data	Scores assigned against descriptors (fluency, complexity, accuracy)	Discourse data (measures of fluency complexity and accuracy)	Questionnaire responses (re task difficulty and task enjoyment)
Analysis	FACETS analysis, <i>t</i> -tests	ANOVA (MANOVA)	<i>t</i> -tests
Findings	No effect	No effect	No effect

Table 3. Impact of planning condition on overall ability estimates (paired sample *t*-test)

Condition	Measure (logits)	Standard Error	<i>t</i>	<i>p</i>
+	-0.04	0.06	0.67	ns
-	+0.04	0.06		

Table 4. Impact of planning condition on fluency, accuracy and complexity measures (paired sample *t*-test)

Criterion	Condition	Mean (raw score)	Standard Error	<i>t</i>	<i>p</i>
Fluency	+	3.47	0.06	0.36	ns
	-	3.49	0.05		
Accuracy	+	3.50	0.05	0.08	ns
	-	3.50	0.06		
Complexity	+	3.63	0.05	0.64	ns
	-	3.65	0.05		

Table 5. Descriptive statistics for fluency, accuracy and complexity discourse measures (*M*, *SD*)

Condition		Fluency				Pauses	Accuracy no of error free c-units	Complexity Clause/ c-unit
		False starts	Hesita- tion	Refor- mulation	Repeti- tion			
+	<i>M</i>	0.01	0.05	0.16	0.2	0.15	67.67	1.4
	<i>SD</i>	0.02	0.06	0.12	0.22	0.17	17.96	-0.24
-	<i>M</i>	0.01	0.05	0.18	0.21	0.17	64.55	1.41
	<i>SD</i>	0.02	0.06	0.12	0.23	0.15	19.13	-0.23

Table 6. Impact of planning condition on fluency discourse measures (Two-way repeated measures MANOVA, Wilks test)

Source	Value	<i>F</i>	Hypoth <i>df</i>	Error <i>df</i>	<i>p</i>	Effect size
<i>Planning</i>						
Version	.86	.97	5.00	30.00	.45	.13
Condition	.77	1.76	5.00	30.00	.15	.23
Condition * Version	.91	.59	5.00	30.00	.71	.09

RQ 2 Does the provision of strategic planning time make a difference to language production as reflected in features of the oral discourse produced by test-takers?

This question was answered by means of a quantitative analysis of test discourse to determine whether measures of accuracy, fluency and complexity in participant output differs according to the hypothesized difficulty of the performance condition. Results (see Tables 5–8) revealed that strategic planning had no effect on performance when analyzed according to any of the above

Table 7. Impact of planning condition on accuracy discourse measure (two-way repeated measures ANOVA)

Source	Sum of squares	<i>df</i>	Mean square	<i>F</i>	<i>p</i>	Effect size
Within subjects effects						
Condition	367.20	1	367.20	2.79	.10	.08
Condition * Version	215.27	1	215.27	1.63	.21	.05
Error (Condition)	4346.02	33				
Between subjects effects						
Version	137.22	1	137.22	.24	.625	.01
Error (Version)	18554.91	33	562.27			

Table 8. Impact of performance conditions on complexity discourse measure for each task dimension (two-way repeated measures ANOVA)

Source	Sum of squares	<i>df</i>	Mean square	<i>F</i>	<i>p</i>	Effect size
Within subjects effects						
Condition	.11	1	.11	2.19	.15	.06
Condition * Version	.00	1	.00	.02	.90	.01
Error (Condition)	1.68	34	.05			
Between subjects effects						
Version	.01	1	.01	.22	.64	.01
Error (Version)	2.08	34	.06			

measures. There was a slightly higher number of pauses, reformulations and repetitions and a slightly lower rate of error free clauses under the no planning condition, but these differences were not statistically significant.

RQ 3 Are differences in the amount of strategic planning time provided associated with differences in test taker attitudes and perceptions of task difficulty?

The results in Table 9 show that in Task One test candidates found telling the story slightly easier when they had three minutes of planning time than when they did not. This difference was not however significant and note that the reverse trend (also non significant) was observed for Task Two. There were no significant differences in levels of enjoyment according to the task condition on either Task One or Task Two.

When responses across the two tasks were combined, perceptions of task difficulty and task enjoyment did not differ significantly between the plus and minus planning conditions. However it should be noted that the trend was in

Table 9. Impact of performance conditions on questionnaire responses

Version		Task difficulty (Q1)				Task enjoyment (Q5)			
		+	-	<i>t</i>	<i>d</i>	+	-	<i>t</i>	<i>d</i>
1	M	2.74	2.65	0.07	0.10	2.34	2.35	0.06	0.01
	SD	0.95	0.91	$p = 0.49$		0.74	0.78	$p = 0.51$	
	<i>n</i>	97	100			96	100		
2	M	2.47	2.52	0.36	0.05	2.80	3.01	1.56	0.22
	SD	0.75	0.82	$p = 0.72$		0.91	0.99	$p = 0.12$	
	<i>n</i>	99	97			99	97		

the predicted direction, i.e., planning was perceived to make the task slightly easier (as was true also for the learners taking the familiar task in the Wigglesworth 2000 study), whereas telling the story without the aid of planning was perceived to be somewhat more enjoyable.

Discussion

The findings of this study offer little support for the notion that the provision of planning time in the context of a tape-based test of speaking proficiency makes a difference to the quality of speech produced by the candidates. Although there was a slight tendency for test takers to perceive the tasks to be easier when strategic planning time was provided, this did not result in their achieving a higher score. While scores may be considered rather crude indicators of proficiency, since they are based on impression alone aided by a rating scale with rather sketchy characterizations of performance, the more objective discourse measures of fluency, accuracy and complexity, which parallel those used in other studies, were also invariant across the two performance conditions. These findings are at odds with those reported in previous research (e.g., Crookes 1989; Foster & Skehan 1996; Wigglesworth 1997; the chapters in Section 3 of this book) and therefore require some discussion. We offer the following tentative explanations, most of which relate to characteristics of the tasks used in this study and the conditions under which they were performed.

1. The tasks used in this study were monologic, and conducted in a language laboratory. The absence of a live interlocutor, may have some bearing on the nature of performance reducing candidates motivation to 'lift their game' or to strive for enhanced performance even when opportunities to do so are provided. If this is the case, however, it has implications for

the validity of semi-direct tape based speaking tests more generally. Recall however that in Wigglesworth's (1997) study, the test tasks were also monologic and significant effects were found in the test discourse for fluency, accuracy and complexity even though a smaller amount of planning time (one minute) was given.

2. The tasks used in the study were simple narratives, and feedback from our raters suggests that these may not have been conducive to eliciting complex performances from candidates perhaps because of the sequential nature of the task which places emphasis on moving forward with the story line rather than elaborating on particulars. Comments from test-takers recorded on the open-ended section of the questionnaire suggest that planning time was seen largely as a means of deciphering the pictures and working out the plot of the story rather than as an opportunity for rehearsing the linguistic features of their production (and this is in keeping with what Wigglesworth (2000) proposed by way of explanation for the null finding in her second study). Further studies exploring planning time with a range of different task-types may produce different findings.
3. The task instructions, "you will have a further three minutes to think about the story", which correspond to those currently provided on the TSE, did not invite any kind of focus on form. It would be interesting to find out whether the type of language L2 learners generate differs according to the instructions they are given and also to their awareness of the criteria used for rating their performance. In addition, it would be valuable to explore in more depth than we have been able to do here what learners actually do when they plan in a test environment (see previous studies in classroom contexts by Ortega (1999), and by Sanguran and Ortega in this volume).
4. The participants in this study may have been unfamiliar with speaking under planned conditions or uninformed about how best to use the time available. Allowing 3 minutes for planning, after all, is not a common feature of speaking tests (see our earlier discussion). The fact that many learners perceived the planning time to be beneficial but failed to produce more fluent and more complex or accurate speech suggests, as proposed above, that whatever action they were taking to improve their performance was ineffective. It might be worth training learners in appropriate strategies for making use of planning time in a test situation to see if this makes a difference to the quality of their performance. An unpublished study by Rutherford (2001) gives grounds for optimism here. Her research offered suggestive evidence that even the simple fact of answering a questionnaire about planning strategies after completing a task, can result in learners us-

ing a large range of strategies and producing more complex discourse on a subsequent task.

5. The length of planning time offered to test-takers was insufficient to produce measurable differences in performance. Mehnert's study, as already noted, showed that 10 minutes of planning were required to produce a more complex spoken performance. However the same could not be said for either fluency or accuracy, which were responsive to only 1 minute of planning time – and this was also the case for some aspects of learner discourse in the Wigglesworth (1997) study. These previous research findings thus suggest that the limited amount of planning time offered to these candidates cannot be the full explanation for our non significant results.
6. The distinction between the + and – planning conditions was not sufficiently marked to yield different kinds of performance among the candidates. This is perhaps a more plausible explanation of our null findings than the limited planning explanation proffered above. The 75 seconds provided before participants embarked on telling the story in the unplanned condition may have been sufficient to allow the necessary advance planning to take place. Related to the above point is the possibility that the time allowed for task completion was too generous. Candidates had as much as three minutes to tell the story and were aware of this before starting to tell the story. They may therefore have been able to engage in a considerable amount of on-line planning under both performance conditions and were unlikely to be experiencing a high degree of communicative pressure. Yuan and Ellis (2003) have demonstrated that on-line planning can increase both the grammatical complexity and accuracy of L2 production, so we may surmise that our candidates were able to monitor and improve the quality of their speech even in the absence of strategic planning time. If this is the case it may be appropriate, in the interests of sampling language ability under different performance conditions, to impose more stringent time pressures on candidates on at least some oral test tasks. However, eliminating task familiarization time altogether might be perceived as unfair in that it would prevent candidates from producing their best possible performance (as advocated by Swain (1985a)).
7. The positioning of the tasks within the larger experiment was not conducive to differences in performance. Recall that the tasks used to investigate the planning dimension were administered somewhat late in the piece (i.e., in 6th or 7th position). Although the overall design of the study was counterbalanced, there may nevertheless have been a practice effect or a fatigue effect coming into play at this late stage in the proceedings. The

practice opportunity may have given candidates a chance to hone their story telling skills (albeit on different picture stories) to a point where they had little need for, or experienced few benefits from planning (and see Bygate (this volume) on the effects of task repetition or rehearsal). The debilitating effect of multiple story-tellings on the other hand may have rendered learners unwilling or unable to make the most of the planning time provided. Tentative support for this speculation is provided by the questionnaire feedback showing that the candidates tended to enjoy the task more when planning time was not given. Perhaps they were keen to get the task over and done with.

While the above factors may have contributed either individually or in combination to the null findings of this study, we must also be prepared to entertain the possibility that the result is due to the fact that language behaviour in a test situation is fundamentally different from language of the classroom or of real world encounters and therefore that findings of SLA classroom research (to the extent that these replicate the processing conditions present in natural speech production) may not be generalizable to testing contexts. As noted in previous reports on this study (Iwashita et al. 2001; Elder et al. 2002) the cognitive focus in a test environment is surely on display, for the purpose of having one's ability assessed, rather than on getting the message across. This inescapable fact may well be paramount in test takers minds and override our best efforts to manipulate task characteristics for the purpose of exploring variations in language proficiency under normal 'real world' performance conditions. If this is the case, it raises a major question about our claim to be able to test language communicatively or, in other words, to mirror the demands of real world language encounters in assessment contexts.

However, before we leap to dismiss the relevance of language test data for SLA research there are further avenues to investigate. There have, for instance, been serious attempts to assess speaking dialogically under conditions which more closely approximate authentic communication than does a solo performance in the presence of a tape-recorder (e.g. in the context of peer-to-peer interaction (Swain 2001)) and the effect of planning time on performance under test conditions such as these has yet to be explored. Furthermore, while the need for accountability and fairness places inevitable constraints on the extent to which test tasks can replicate real world processing conditions, recall that at least one test-based study has found a significant effect for planning time on the basis of discourse measures which are the stock tools of trade for SLA research. Intriguing questions remain as to why findings in this area are inconsistent and

further investigations of planning time in testing contexts are certainly worth undertaking both in the interests of understanding more about speaking proficiency per se and also with the aim of producing richer tests which allow inferences to be drawn about candidate's ability in a variety of performance situations. The parameters of this research agenda have been foreshadowed in the discussion above and are summarized below.

Experimental studies of planning time in language testing contexts will need to pay attention to the multiple characteristics of the task (including the important issue of what kind of pre-task stimulus is provided and what instructions are offered to candidates); to test modality and register (i.e., whether candidates are talking to a machine or to one or more 'live' interlocutors of the same or different status); to issues of affect including learners' attitude and commitment to doing their best on the task in question; to the amount of pre-task preparation time and on-line planning opportunities available; to any prior instruction or training which test-takers have received in planning techniques; to what test-takers do with their planning opportunities both before and during their speech performance and, in addition, to the nature of assessment criteria and whether test-takers' attention is drawn to these criteria before embarking on the task. Finally, since research oriented to better test design must, for practical reasons, rely on scores derived from ratings rather than on discourse measures, consideration must be given to the as yet unexplored issue of raters' response to performances which they know to have been produced with and without planning time and the extent to which this awareness results in unwitting bias or deliberate compensatory rating behaviour in relation to one or other condition.

Notes

1. Structure was operationalized in terms of the amount of information provided to the learners to assist them doing the task. Familiarity was operationalized according to whether the task activity was an activity with which the learners would reasonably be expected to be familiar (Wigglesworth 2000: 86).
2. The Next Generation TOEFL test is scheduled to be launched in 2005.
3. Most students lived with Australian host-families, and so unless their host families speak the students' native language, they had to speak English at home.
4. A c-unit was defined as a simple clause, or an independent subclausal unit, together with the subordinate clauses associated with them (Forster, Tonkyn & Wigglesworth 2000).

Appendix

Rating scales

Fluency

- 5 Speaks without hesitation; speech is generally of a speed similar to a native speaker
- 4 Speaks fairly fluently with only occasional hesitation, false starts and modification of attempted utterance. Speech is only slightly slower than that of a native speaker.
- 3 Speaks more slowly than a native speaker due to hesitations and word-finding delays
- 2 A marked degree of hesitation due to word-finding delays or inability to phrase utterances easily
- 1 Speech is quite disfluent due to frequent and lengthy hesitations or false starts

Accuracy

- 5 Errors are barely noticeable
- 4 Errors are not unusual, but rarely major
- 3 Manages most common forms, with occasional errors; major errors present
- 2 Limited linguistic control: major errors frequent
- 1 Clear lack of linguistic control even of basic forms

Complexity

- 5 Confidently attempts a variety of verb forms (e.g. passives, modals, tense and aspect), even if the use is not always correct. Regularly takes risks grammatically in the service of expressing complex meaning. Routinely attempts the use of coordination and subordination to convey ideas that cannot be expressed in a single clause, even if the result is occasionally awkward or incorrect.
- 4 Attempts a variety of verb forms (e.g. passives, modals, tense and aspect), even if the use is not always correct. Takes risks grammatically in the service of expressing complex meaning. Regularly attempts the use of coordination and subordination to convey ideas that cannot be expressed in a single clause, even if the result is awkward or incorrect.
- 3 Mostly relies on simple verb forms, with some attempt to use a greater variety of forms (e.g. passives, modals, more varied tense and aspect). Some attempt to use coordination and subordination to convey ideas that cannot be expressed in a single clause.
- 2 Produces numerous sentence fragments in a predictable set of simple clause structures. If coordination and/or subordination are attempted to express more complex clause relations, this is hesitant and done with difficulty.
- 1 Produces mostly sentence fragments and simple phrases. Little attempt to use any grammatical means to connect ideas across clauses.

Appendix A

Story 7

Dog, car and ladder (Planning +)

INSTRUCTIONS

Now please look at the six pictures on the right.

I would like you to tell the story that the pictures show, starting with picture number 1 and going through to picture number 6.

Please take 45 seconds to look at the pictures, and then you will have a further 3 minutes to think about the story.

You may take notes on the sheet provided.

Do not begin the story until you are told to do so.

You will have approximately 3 minutes to complete the story.

You will not be allowed to look back at your notes.

Please read out the title of the story "Dog, car and ladder" before you begin.



Figure A.

CHAPTER 9

Strategic planning, task structure, and performance testing*

Parvaneh Tavakoli and Peter Skehan

Kings College, London / Chinese University of Hong Kong

Introduction

Pollitt (1990) makes a distinction between approaches to language testing which *count*, e.g. a multiple choice reading test, and those which *judge*, e.g. with rating scales. A “judging” approach to assessment is a form of performance testing. Within this area, there has been significant interest in recent years in the use of tasks to underpin the assessment decisions that are made. There are several reasons for this. First, task-based research has been an active area within second language acquisition, and as a result, there are a range of research findings which could, potentially, be related to language testing contexts. Second, there are the beginnings of models of task-based performance which could be the basis for applications to assessment. Above all, though, researchers in this neighbouring area have attempted to explore how tasks can be described and analysed, and their properties linked to the nature of the performance that is elicited. To put this another way, researchers generally argue that tasks are not neutral devices which elicit performance in a straightforward manner – they influence the nature of the performance which results. As a result, insights from this closely related area may clarify how performance-testing procedures may need to take account of the systematic influences, as well as level of difficulty, that tasks provide. For these reasons, we will next review some of the recent task research.

Table 1. Task influence on performance dimensions

Task characteristic	Accuracy	Complexity	Fluency
familiarity of information	greater	no effect	slightly greater
dialogic vs. monologic tasks	greater	greater	lower
degree of structure	greater	no effect	greater
complexity of outcome	no effect	greater	no effect
transformations	no effect	planned condition generates greater complexity	no effect

Modelling and researching task-based performance

Two approaches will be outlined here, following work by Skehan and Robinson. Both approaches are of a generally cognitive orientation, although they differ in the claims they make about the systematic influences of tasks upon performance. Following extensive work within cognitive psychology (see, e.g., Miyake & Shah 1999), Skehan (1996a, 1998b) makes the assumption that human beings operate with limited capacity attentional systems, and that to pay attention to one area of performance may well be to reduce the attention available elsewhere. In other words, if performance is multi-dimensional, improving performance in one area may well cause achievement in other areas to be lowered. In a series of studies, Skehan and Foster and their collaborators have demonstrated that a number of task characteristics have systematic influences upon performance. First of all, regarding performance itself, they have shown (Skehan & Foster 1997) that it is useful to explore the complexity of language, its accuracy, and its fluency, and that these three areas enter into competition with one another for scarce attentional resources. Second, the range of evidence from several studies is consistent with different influences upon each of these performance areas. This is shown in the following table (taken from Skehan (2001)).

The claim here is that a set of task characteristics have predictable influences upon performance. For example, if a task draws upon familiar information, then, other things being equal, it is likely to yield a performance which is more accurate and more fluent, but without any particular impact upon the complexity of the language which is used. If one takes into account that raters

of performance will be attending to areas such as accuracy and fluency, then the use of familiar information tasks is likely to give such raters reasons to grade more positively. If all tasks were based on such information, this would not be an issue, but if tasks varied, without control, such that different test takers were rated on tasks which varied in the familiarity of information they were based on, the differences in the ratings given could be partly artifactual. The same issue applies to the other task characteristics that influence performance.

A more problematic area within this research has been to explore task characteristics which impact upon the general construct of task difficulty. Skehan (1998a), for example, proposes that a number of features impact upon difficulty, including:

- number of elements in the task (more means more difficult);
- type of information (concrete (less difficult) vs. abstract)
(Brown et al. 1984).

But there are difficulties with difficulty, and these difficulties become more salient when task-based performance testing is being considered. A concept such as difficulty implies an overarching dimension into which other features fit. But we have seen that task performance is multidimensional, so that devoting attention to one area may be at the expense of others. In this case, the problem is to decide on an *overall* level of difficulty for a task. Would it be the general complexity of language achieved? Or the accuracy? These difficulties currently seem to have no obvious solution for testers, and so using tasks to assess the capacity to speak contrasts with the degree of success that has been achieved in predicting difficulty in other areas, e.g. reading (Kirsch & Mosenthaler 1990).

Finally, in this review of task research, it is clear that the conditions under which a task is done can exert a strong influence upon performance. The bulk of the research here has been into the effects of pre-task planning (strategic or rehearsal). It is now clear that giving learners (or test-takers) such planning time leads to significant improvements in performance. The clearest generalisation is that pre-task planning is associated with greater complexity and fluency, in almost all studies. There is a less robust relationship with accuracy, with some studies supportive of an accuracy effect, and others much less so. Further work is needed to uncover which particular conditions in planning lead to greater accuracy.

As a result of this range of studies, Skehan (1998a, 2001) proposes a model of task-based performance in relation to language testing (see Figure 1).

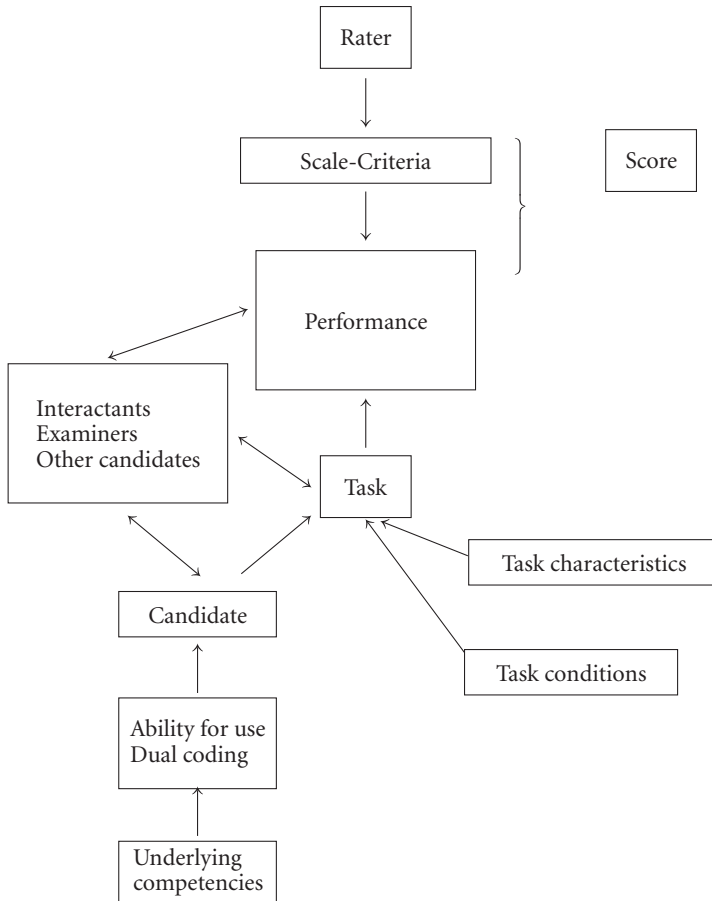


Figure 1. Task based performance and language testing

The broadest purpose of this model is to make clear that the rating assigned someone on the basis of their performance on a task is the consequence of a whole range of factors, only one of which can possibly be their underlying competence. In addition, we have to consider:

- the method by which the rating is done, with the potential this has to introduce error;
- the context for the performance, including the nature of the interactants involved, and their relationship to one another;
- the extent to which the testee can engage strategies of performance, and general processing skills, handling rule-based and memory-based language;

- the task that is involved and the conditions under which it is done.

Essentially, these different factors, besides helping us to understand how test scores may be the result of a multiplicity of influences, also provide an agenda for research. With regard to the study to be reported below, this means that we need to advance our understanding of the influence of task characteristics on performance, as well as what impact the conditions under which tasks are done might have on that same performance.

There is, though, an alternative perspective on the use of tasks, and of their potential for assessment, which is represented in the work of Peter Robinson, (see Chapter 1). Three main differences are apparent. First, Robinson (2001c) does not assume that attention is limited in capacity, preferring instead to view attention as an expandable resource, particularly in relation to memory in different, non-competing modality areas. Second, Robinson sees tasks as driving performance in a slightly different way. For him, language complexity follows functional complexity (Givon 1985). In other words, he predicts that more difficult tasks will push learners to engage more complex language *and* push them to achieve accuracy so that communication is more effective. In other words, he is proposing that learners will “rise to the challenge” of completing more difficult tasks, and perform better. Third, Robinson distinguishes between what he terms resource directing and resource depleting factors. The former are the influences which push learners to engage in more difficult language and are therefore what task designers need to attend to if they are to influence learner (and test-taker) performance. The latter, resource depleting factors, such as the opportunity for pre-task planning time, are those which influence the overall difficulty of a task.

The relevance of task-based studies within an SLA performance framework has certainly been noticed within language testing. Iwashita et al. (2001) used task features identified by Skehan (1998b, 2001) and Robinson (2001c) and explored the relevance of these features for test-task performance. They focussed on a narrative task, on the basis that this task-type is the most typical of testing formats. This is because it is non-interactive, and so potentially open to greater control, and therefore standardisation, on the part of the test producer. In particular, they explored the effects of:

- *perspective*: whether a narrative story is told from the teller’s point of view, or from someone else’s point of view;
- *immediacy*: whether the test taker told the story with the pictures in front of them, or without the pictures present;

- *adequacy*: whether test takers told the story with access to all the pictures in a picture series, or with two of the pictures (out of six) missing;
- *strategic planning time*: whether there was three minutes planning time or not.

The results obtained did not follow the predictions arising from the task literature. Only one of the twenty-four comparisons yielded a significant result. This was for the immediacy condition, with accuracy scores. All other comparisons were non-significant. Worse, the one significant result was *against* the predicted direction, with the more difficult non-immediate conditions producing greater accuracy.¹ Clearly, these results are disappointing from a task research perspective, and lead to three alternative interpretations. First, it may be that the narrowness of the range of tasks used (all picture-prompt narratives), and the constraints of testing conditions lead to the failure to find any significant differences. Second, it is possible that the variables which were manipulated were not the most salient ones from the task literature. Third, even if the variables manipulated were well chosen, it is possible that their operationalisations were not ideal. But these three interpretations are all post-hoc attempts at rationalising what one would have expected to have been a series of significant results. Obviously, there is a fourth interpretation – that task manipulations are relevant to pedagogy and certain types of performance, but not to the context of testing. The study by Iwashita et al. (2001) presents a challenge to those who think that the task-based literature can make contributions to assessment.

Task structure and task performance

Further work is clearly required to help clarify the precise ways in which task research may be relevant for assessment and the ways it may not. In Table 1, five task characteristics were proposed which may impact upon performance: information familiarity, dialogic vs. monologic tasks; degree of structure; complexity of outcome; and number of transformations. The first of these was included by Iwashita et al. (2001) and shown to have only a marginal influence on performance. The second could be regarded as difficult to implement in testing, to the extent that there is a need to have standardised conditions, of a sort that would be compromised with multiple participants in a task. It could also be argued that Iwashita et al. (2001) indirectly explored the feature, ‘transformations’, when they examined ‘adequacy’, in that missing pictures in

the task they used required test takers to transform the input material. The remaining two features would seem worth investigating. In the present study, the focus will be on the first of these, degree of structure, principally because this task feature is more readily manipulable with narrative tasks.

In two early studies, Foster and Skehan (1996) and Skehan and Foster (1997) explored three tasks: a personal information exchange task, a narrative, and a decision-making task. Predictions were made about the effect of task type on task performance, and these predictions were partially borne out. However, there were aspects of performance which did not follow the predictions. A *post-hoc* interpretation of these studies suggested that the tasks which contained a clear macrostructure seemed to advantage fluency and accuracy, but left complexity unaffected.

Given the *post-hoc* nature of these interpretations, a subsequent study was designed with the intention of comparing tasks containing structure with those which did not. Participants were required to retell narratives based on video prompts. The prompts were, in fact, Mr. Bean episodes. One of these “Crazy Golf”, involved Mr. Bean playing a round of crazy golf. Over-interpreting the golf attendant’s instruction that he should, on no account, touch the ball while playing a hole, Mr. Bean ended up knocking the golf ball outside the golf range which led to a series of unlinked events. There was no structure to these events, and the way one event moved to the next was fairly unpredictable. The second video prompt was “Mr. Bean goes to the restaurant”, reflecting Mr. Bean treating himself to a meal on his birthday. Although a number of things happened which do not normally happen in restaurants, e.g. Mr. Bean hiding a disgusting meal in the pockets of fellow diners, the general structure of the narrative simply followed the conventional restaurant “script”: i.e. greeting by maitre d’hotel, being seated, being shown the menu etc. Results were supportive of the predictions. The structured narrative (Restaurant) produced by the participants was considerably more fluent than the unstructured narrative, and, when there was strategic planning, it was also more accurate.

The results of this study are encouraging for our understanding of the role of task structure on performance. However, they do raise some additional questions. Clearest of these is that it is important to make progress in understanding what task structure is, and how it can be characterised. Foster and Skehan (1996) and Skehan and Foster (1997) represented task structure as consisting of a clear time line, so that “structure” was introduced into the task as the clarity of the macrostructure involved when a series of events unfolded in time. It was assumed that the second language speakers clearly apprehended an overall macrostructure and so were aware of this macrostructure in general while con-

centrating on telling particular parts of the story. To portray this in relation to Levelt's model of speech production (Levelt 1989), it is hypothesised that the Conceptualiser component of the storytelling is relatively unpressured, with the result that attention can be more easily allocated to the Formulator, and this will impact upon fluency and accuracy. In the Mr. Bean restaurant task, similarly, it was assumed that general knowledge of a restaurant "script" would lead learners to have attention available to achieve greater accuracy and fluency with the details of the story to be told.

The difficulty here is that while one may have a general notion of a macrostructure in the speaker's mind, how that macrostructure is justified or described is not so clear. Already we have:

- a clear time line
- a script
- a story with a conventional beginning, middle and end
- an appeal to what is familiar and organised in the speaker's mind

with all of these functioning to create a macrostructure which will then impact upon performance. Essentially, this raises the possibility that macrostructures can be created from a number of sources, and that we therefore need to make progress in understanding what the possible sources are.

In fact, the four characteristics just mentioned do not exhaust the different ways in which a task can be regarded as structured. Other work in testing may be instructive. Kobayashi (1995, 2002), following work by Winter (1976) and Hoey (1983), demonstrated the relevance of a problem-solution structure, i.e. a narrative sequence whose centre is a problem which is resolved. Optionally this structure can have the more extended sequence *Situation* (which sets the scene and introduces relevant information) – *Problem-Solution* and *Evaluation*, (which functions as a sort of commentary on the satisfactoriness (or otherwise) of the proposed solution). Kobayashi (1995, 2002) showed that reading comprehension texts based on a problem-solution structure produced different results to those which were not structured in this way. The structured texts distinguished more clearly amongst *more* proficient students, especially when comprehension was measured by summary and open-ended questions, rather than a cloze test format.

It follows that a clear macrostructure can be achieved by ensuring logical relations between the elements of the story instead of relying on a clear and simple story line or on some sequence of events familiar to the learners. In a sense therefore there may be a greater degree of universality to the structure involved, since it will be based on causal connections. Such a set of links

might be regarded as having a “tighter” organising frame than other sorts of macrostructure.

This analysis suggests an interesting possibility: that within the general concept of macrostructure, there are different varieties of macrostructure available, and that these different varieties may be arrangeable in the form of a cline, reflecting the *degree* of structure. This would lead to the prediction that if task structure leads to greater fluency and accuracy, then greater task structure would lead to particularly high fluency and accuracy. This would be interesting from a pedagogic or a testing perspective. It could also be the basis for research, since it would generate measurable comparisons between tasks at different levels of structure.

Another area which impacts upon task and test performance is that of strategic planning. A number of studies have shown clear effects in this area, effects which are reviewed in the chapters of this book. In general, the results have shown that giving learners planning time has a favourable effect upon performance, with consistent and appreciable increases in complexity and fluency (Skehan & Foster 2001) and less consistent, and smaller increases for accuracy. The fluency effects do seem the clearest and most consistent. With complexity (which connects with the concept of *range* in the testing literature) most studies report gains, but there are some contradictions. Wigglesworth (2001), for example, working in a testing context, does not find the usual effects here, raising the possibility that the prominence of an assessment framework may cause differences in the results which are found. Regarding accuracy, a series of studies have shown significant effects, such as Foster and Skehan (1996, 1999), Mehnert (1998) and Skehan and Foster (1997, 1999). But other studies have not reported similar outcomes (Crookes 1989; Ortega 1999; Wigglesworth 2001). In this latter case, there were slightly different patterns of results at two different proficiency levels, suggesting that the effects of the independent variables may interact with proficiency level. It is also possible that there is an interaction between strategic planning, performance area and task type, since the tasks used by these different investigators have varied. There may be a tendency, in fact, for narrative tasks, the very tasks most favoured in testing contexts, to be least likely to deliver a significant accuracy effect. Although there are studies using narrative tasks in which accuracy effects are found (Ellis 1987; Yuan & Ellis 2003), these only use narrative tasks, and so a comparative task dimension is lacking. Where studies use more than one task type, e.g. Foster and Skehan (1996), Skehan and Foster (1997), the narrative task is associated with smaller accuracy effects than, for example, decision-making and personal information

exchange tasks. In any case, it is clear that there is further scope to investigate the effects of strategic planning in this particular performance domain.

Drawing on this review of the literature, with tasks and with planning, a number of hypotheses can be formulated:

Hypothesis One: degree of structure in a narrative task will influence the fluency and accuracy of performance on the task. Further, progressively greater structure, in the sequence no structure>schematic structure>weak causation structure>problem solution structure will lead to greater increases in fluency and accuracy.

Hypothesis Two: there will be no influence of task structure on complexity.

Hypothesis Three: strategic planning will influence complexity, fluency and accuracy.

Hypothesis Four: the effects of these variables will not vary as a function of proficiency level.

Method

Design

A $2 \times 2 \times 4$ factorial design was used in the current study with pre-task planning condition, proficiency level, and task structure as the independent variables. Planning condition and language proficiency were between-participants variables and each had two levels with the participants belonging to either of the two conditions and levels. Task structure, which was operationalised through 4 different picture series, had four levels representing a scale in the degree of structure of a task. Task structure was a within-participant variable, i.e. all participants performed all four tasks.

Task

As discussed earlier, narrative tasks are frequently employed in the context of assessing language performance (Elder et al. 2002; Iwashita et al. 2001; Robinson 2001c). Narrative tasks are also routinely used as a single type of stimulus by some international testing organisations (e.g.: Test of Spoken English). Narrative tasks in this sense refer to stories based on a sequenced set of picture prompts, which are given to participants in order to elicit language perfor-

mance. The rationale for using narratives is justified in terms of construct validity, reliability and authenticity of the test. However, the prime reason for selecting narratives in the present study is to have conformity with the literature from which the theoretical assumptions of the study are drawn.

In order to find appropriate picture series, two main sources were consulted:

1) EFL sources including course books and supplementary materials for teaching English and other modern languages; and 2) non-EFL sources including a wide range of different materials such as cartoon books, newspapers and pictorial stories. The aim was to find picture series which were clear, had worthwhile stories to be told, were of a length suitable for the study, were culturally familiar to the participants, were neither linguistically cued nor linguistically demanding, and looked interesting. The picture series identified in this way were then carefully examined and categorised into structured and unstructured. Structured tasks were defined as having either a problem-solution structure or a schematic sequential structure. The unstructured picture series were further distinguished between completely unstructured series and those containing minimal structure. Lack of structure was operationalised in terms of the number of pictures, other than the first and the last, that could be interchanged with one another without the story being compromised.

To achieve the purpose of the study regarding degrees of structure, two structured tasks and two unstructured tasks were selected. From among the structured-tasks category, one picture series was selected to represent the problem-solution structure and one picture series was selected to represent the a weak causation structure. Based on the theoretical assumptions of structure discussed earlier, the two structured tasks differed from one another in terms of the type of structure they exposed and the degree of structure they presented. In fact, following Kobayashi (1995, 2002), the problem-solution structure was assumed to have a stronger type of structure than the weak causation structure because the solution involved has a greater unifying or “resolving” effect on the overall story macrostructure. The task selected in the problem-solution category, i.e. the Football task, was a picture series with a transparent problem-solution structure and a well-presented sequential organisation. The second structured task, the Picnic task, on the other hand, was based on a clear organisation and contained an implicitly stated problem, which was only revealed in its last frame. However, this task did not propose a clear transparent problem-solution structure, which made it less structured than the Football task. Therefore, both tasks were structured but differed in the degree of structure they demonstrated.

Similarly, two tasks with varying degrees of structure were selected from the unstructured group. The lack of a problem-solution relationship on the one hand and lack of causative elements on the other hand suggested that both tasks were unstructured. However, they differed from one another with regard to the amount of sequential organisation they contained. Task four, the Walkman task, did not contain any sequential organisation and, therefore, was less structured than task three, the Unlucky Man, which had a loosely presented sequential organisation. In effect, events in task four were loosely related to one another and the sequence of organisation of events hardly followed a timeline. The four picture sets can be found in Appendix 1.

All the picture sets consisted of six pictures, except the Unlucky Man task, which due to the assumptions of the study, had a set of moveable pictures in the middle and, therefore, had ten pictures. Figure 2 indicates how the four tasks can be located on a continuum representing a scale of the degree of structure hypothesized in the study reported here:

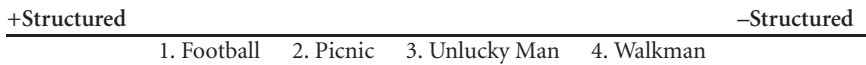


Figure 2. Degrees of structure in the four tasks

In order to avoid a practice effect, a counterbalanced design was used. In effect, participants performed all four tasks but in four different sequences. Table 2 demonstrates the four sequences of tasks in the study.

Table 2. Counterbalanced sequence of the tasks

Sequence 1	Football	Picnic	Unlucky Man	Walkman
Sequence 2	Picnic	Unlucky Man	Walkman	Football
Sequence 3	Unlucky Man	Walkman	Football	Picnic
Sequence 4	Walkman	Football	Picnic	Unlucky Man

Strategic Planning Conditions: As planning was a between-participant variable, half of the participants performed the tasks under planned and half under the unplanned conditions. Planning was operationalised in terms of the amount of strategic planning time provided to the participants. As discussed earlier, the amount of time given to the participants was determined on the

basis of the findings of previous studies (Elder et al. 2002; Mehnert 1998; Wigglesworth 2001). The unplanned group was given 30 seconds to look at each of the picture series before they started telling the stories. Participants in the planned groups were given 5 minutes to look at each of the picture series and were advised to plan for what they were going to say. Moreover, under the planned conditions each participant was given a sheet of paper to take notes. However, they were informed that they would not be allowed to use their notes while they were telling the story. The instructions given to both groups were identical in all other respects.

Language proficiency level: The participants were drawn from two levels of language proficiency, i.e. elementary and intermediate. Prior to the study the candidates were placed in to their levels on the basis of an institutional placement test. Nevertheless, to confirm the homogeneity of the groups and also the distinction between the two proficiency levels, their language proficiency was tested by the “Oxford Placement Test 2” (Allan 1992). It should be noted that due to the practical limitations, only the grammar part of the test was run. Participants’ responses were checked and scored on a scale of 100 points. The elementary group had a range of scores between 17 to 44 and the intermediate group scored between 45 and 75. These grammar-test results confirm the decisions made on the basis of the wider-ranging institutional placement test.

Perceptions of task difficulty: In order to explore participant perceptions of task difficulty, separate (but related) questionnaires were developed for each of the planning conditions. All participants were asked to complete the appropriate questionnaire as soon as they performed the four tasks. Both questionnaires contained questions about participant perceptions of task difficulty as well as an open-ended question in which participants could give suggestions and comments about the tasks. The planned-group questionnaire differed from the unplanned-group since it included an extra question about the usefulness of the strategic planning time for each of the four tasks. Regarding task difficulty, answers were given on a four-point scale with 1 representing “very easy tasks” and 4 “very difficult tasks”. The answers to the extra question for the planned group were also given on a four-point scale with 1 indicating that the strategic planning time “helped very much” and 4 showing that the strategic planning time “did not help at all”. To avoid any potential confusion or misinterpretations resulting from the participants’ reading ability, the questionnaires were translated into the participants’ first language.

Pilot study: In order to see whether the selected tasks were functioning in line with the theoretical assumptions of the study, the four tasks were piloted twice. In the first pilot study, three Farsi-speakers, one at elementary and two intermediate, performed the four tasks and completed the questionnaires, with this data revealing one of the tasks to be confusing to participants. This task was then replaced with another task from within the same category of structure. The new set of four tasks and the two planning conditions were then piloted on 14 elementary and intermediate learners in a college in London. The participants were aged between 18 and 24, and were from 3 different language backgrounds (Farsi, Chinese and Arabic). They were assigned to either a planned or an unplanned group and completed the tasks in a one-on-one setting with the researcher. They performed the four tasks in a counterbalanced design and completed the questionnaires afterwards. The results of the pilot study suggested that the selection of tasks and the amount of pre-task planning time were appropriate and practical.

Participants in the main study: Participants in the main study were 80 language learners studying English at an educational association in Tehran, Iran. They were all adult females aged between 18 and 45. They were studying English as a foreign language at an elementary or intermediate level and had been studying English at the same language school for at least 18 months. The participants were all Farsi speakers and had a similar language learning history both in the state school system and at the above-mentioned language school. But they differed regarding the period of time they had been studying English in the past, the contact they had with English outside classroom, and the purposes for which they were studying English.

As they had already taken part in similar testing situations in their language school and performed similar tasks, they were all familiar with both the testing conditions and the test format, i.e. narratives. One participant was withdrawn from the study and replaced as she expressed unwillingness during the test. The participants were randomly assigned to a planned or unplanned condition and one of the four sequences of the counterbalanced design, as demonstrated in Table 3.

Setting: All participants were tested in a one-to-one setting by the first author who met them individually and explained the purpose of the test to them. After each participant was randomly assigned to either the unplanned or planned conditions and to one of the four counterbalanced sequences of the four tasks, the instructions were given to them. The participants were given each of the

Table 3. Design of the study

Planning condition	Proficiency level	No. of participants in sequence 1	No. of participants in sequence 2	No. of participants in sequence 3	No. of participants in sequence 4
Planned	Low-proficiency	5	5	5	5
	High-proficiency	5	5	5	5
Unplanned	Low-proficiency	5	5	5	5
	High-proficiency	5	5	5	5

picture series, in turn, and asked to tell the story to the researcher in a way that she could understand what was happening in each story. Under the unplanned condition, they were told that they had just 30 seconds to look at the pictures before they started telling the story. For each task they were given 3 to 4 minutes to tell the story. After the initial 30 seconds, the participants had the picture series in hand, looked at them and told the story to the researcher who tape-recorded the participant's performance on the first task. Then, the same process was repeated for the second, third and fourth tasks, one after the other. Under the planned condition, the participants were told that they had 5 minutes to look at each picture series and plan what to say, and that they would eventually have 3 to 4 minutes to tell the story. Each of these time intervals was chosen to ensure comparability with task-based studies in the literature with an assessment focus (e.g. Wigglesworth 2001; Iwashita et al. 2001). They were also given some paper to take notes if they wished. They were reminded that they would not be allowed to use their notes while they were telling the stories. After the five minutes, the participants told the story to the researcher who tape-recorded the participant's performance on the first task. Then, the same process was repeated for the second, third and fourth tasks one after the other.

Based on the type of the planning condition, the participants were asked to complete the appropriate questionnaire. They were also encouraged to comment about the test and the tasks in general in the last question of the questionnaire. All the introductory talk and instructions to the participants were given in Farsi.

Analytic measures

The recorded performances of all 80 participants of the study were transcribed, word-processed, and then digitised. The following sections will provide a de-

tailed description of the dependent variables as well as how the speech samples were coded and analysed.

Fluency: Koponen and Riggenbach (2000) have discussed different aspects and representations of fluency in detail and argue that fluency may refer to smoothness of speech in terms of temporal, phonetic, and acoustic features; it may represent proficiency at a macro or micro level; it may mean the automaticity of psychological processes; or it may be expressed as a notion contrasting with the concept of accuracy. Freed (2000) proposes that fluency spans a continuum that ranges from studies of its psychological manifestations and reflections of underlying speech-planning and thinking processes to studies of speech production, hesitation phenomena, and the temporal dimensions of speech.

Based on this multifaceted nature of fluency, different researchers have adopted various measures to assess fluency. These measures, however, can be categorised into some sub-dimensions. The first sub-dimension of fluency is silence, or as Skehan (2003) puts it, breakdown fluency. Length and number of unfilled pauses, filled pauses and total amount of silence are some of the measures researchers have used to assess this aspect of fluency. There is, though, some disagreement regarding the minimum length for a pause to be counted as a pause, with proposals as low as .25 of a second (Kormos & Denes 2004). Freed (2000), in a study aimed at exploring the construct of fluency in the speech of L2 learners of French, investigated fluency in terms of 7 measures including unfilled pauses. Regarding the unfilled pauses, she measured the disfluent-sounding silences occurred at places other than predictable juncture boundaries which tended to be of .4 a second or longer in duration. She argues that:

Since silent pauses of shorter duration, frequently termed micropauses and measured in milliseconds, are characteristic of native speech and accurately measured by computerized acoustic analysis, we chose to identify and measure only those unfilled pauses [.4 a second or larger] that were heard as dysfluent.
(Freed 2000:248)

We will follow Freed's proposals for minimum length of pause in the present research. A second sub-dimension of fluency deals with the speed with which language is produced. Measures of speech rate, articulation rate, amount of speech, time ratio and mean length of run are typical here. Speech rate and length of run are the two most commonly used measures in SLA studies. Mehnert (1998), Towell et al. (1996) and Freed (2000) have used mean length of run to measure fluency of the speech production. Mean length of run in Towell et

al. (1996) is calculated as the mean number of syllables produced in utterances between pauses of .28 seconds and above. Mehnert (1998) found mean length of run by calculating the mean number of the syllables between pauses of 1 second. Freed (2000) defines length of run as continuous streams of running speech (measured in words) not interrupted by disfluent pauses or hesitations. Therefore, mean length of run is a manifestation of how lengthy the language produced between two pause boundaries is. Speech rate, i.e. number of syllables or words on average per minute, is another measure frequently used by researchers as an index of fluency (Yuan & Ellis 2003; Mehnert 1998; Raupach 1980; Robinson 2001c). Freed (2000) measured speech rate on the basis of the number of “nonrepeated” words or semantic units per minute. Towell et al. (1996) have calculated speech rate by dividing the total number of syllables produced in a given speech sample by the amount of total time including the pauses. It can be concluded thus that speech rate refers to how fast and dense the produced language is in terms of the time units.

The third sub-dimension of fluency is what is known as repair fluency (Skehan 2003). Repair fluency includes reformulation, replacement, false starts and repetition of words or phrases. Wigglesworth (1997) measured the percentages of clauses containing self-repairs and reported that planned performance is significantly more fluent than unplanned. Skehan and Foster (1999) used repetitions, false starts, reformulations and replacements to measure fluency in their study of the effect of structure on narrative task performance. Freed (2000) operationalised repair fluency in terms of repetition of exact words, syllables or phrases, reformulations, false starts, corrections and partial repeats in the learner speech.

These various conceptualisations of the nature of fluency have rarely been investigated together in task-based studies. Hence, in order to have a more detailed and precise exploration of the nature of fluency and to know what effects different task characteristics would have on various aspects of fluency in task-based contexts, a wide range of different measures is used, i.e. the number of false starts, reformulations, replacements, repetitions, length of run, speech rate, number of pauses, mean length of pauses and total amount of silence.

Accuracy: With measures of accuracy, there is greater consensus among researchers. In a few studies accuracy has been measured by specific measures, such as past tense morphemes (Ellis 1987) and plural -s (Crookes 1989; Wigglesworth 1997; Ortega 1999). Some of these studies did not reveal any significant differences between different planning or task conditions, e.g. Crookes and Ortega. On the basis of these results, Skehan and Foster (1999) argued

that such specific measures are less sensitive to detecting differences between experimental conditions. As a result, they have used general measures of accuracy, such as the number of error free clauses divided by the total number of clauses (Foster and Skehan (1996) and Skehan and Foster (1999)). In both of these studies accuracy effects were detected when pre-task planning time was provided to learners and when task structure was present. Furthermore, interactions were found between task structure and pre-task planning time. In contrast, Ortega (1999) measured accuracy by means of targetlike use of analysis of two grammatical areas: morphology agreement of a noun and its modifiers (including possessives, adjectives and quantifiers), and use of the Spanish article system. She argues that general measures have the disadvantage of being too broad to capture small changes in targetlike use since they combine multiple error types and obscure errors that might be important at a given level of development. Interestingly, and with a degree of compromise, Mehnert (1998) used general measures (percentage of error-free clauses and the number of errors per 100 words) as well as more specific measures (word order and lexical choice). The results of her study showed that both of the general measures of accuracy generated significance, but that neither of the specific accuracy measures did. This result is consistent with the view that while general measures are more blunt instruments, they do capture more variance in performance, and as a result, are more sensitive to the detection of significant effects.

In the current study, accuracy was measured by an index of error-free clauses. Error-free clauses were defined as clauses in which no error was seen with regard to syntax, morphology, native-like lexical choice or word order. However, errors in stress, intonation patterns or pronunciation of the words and utterances were not included. The native-like use of the language, in terms of the grammar and lexis, was generally considered as a criterion in determining whether the clauses were error-free. All error-free clauses were then identified and coded in the transcribed data, and the ratio of the error-free clauses to the total number of clauses was calculated.

Complexity: Foster, Tonkyn and Wigglesworth (2000) have discussed the analysis of spoken data in detail and emphasised that such analysis requires a principled way of dividing the transcribed data into units in order to assess features such as accuracy and complexity. Identifying the shortcomings of measures like T-units and C-units, they have introduced the AS-Unit, (Analysis of Speech Unit). They provide a number of reasons to show that the AS-unit is more appropriate than units used by previous researchers. Foster et al. (2000) define the AS-Unit as “a single speaker’s utterance consisting of *an independent*

clause, or sub-clausal unit, together with any *subordinate clause(s)* associated with either” (p. 365). In this definition, an independent clause will be minimally a clause including a finite verb. An independent sub-clausal unit will consist of either one or more phrases which can be elaborated to a full clause by means of recovery of ellipted elements from the context of the discourse or situation. The definition of AS-unit also considers minor utterances which are one class of “Irregular sentences” or “Nonsentences” identified by Quirk, Greenbaum, Leech, and Svartvik (1985). Furthermore, Foster et al. (2000) explain that “a subordinate clause will consist minimally of a finite or non-finite verb element plus at least one other clause element (Subject, Object, Complement or Adverbial)” (p. 366).

Following Foster et al. (2000), the transcribed data was coded into AS units that contained independent clauses, subordinate clauses and sub-clausal units. The intonation and pausing patterns of speech had a direct influence on determining whether a clause was an independent clause or a dependent one. As a result, the complexity of the performance was measured through an index of subordination by dividing the number of clauses by the number of AS units.

Coding the data and inter-rater reliability. Once the data were coded, 10% of the data was coded by an independent expert against which the data coded by the first author was tested. The inter-rater reliability coefficients were all above 0.90 for the codings of the complexity measures, i.e. the AS units and the dependent clauses, as well as repetitions and replacements. However, the reliability coefficient for measures of accuracy, false starts and reformulations were initially lower. As a result, there was a reassessment of the measures of accuracy, false starts and reformulations until inter-rater correlations of above .94 were achieved.

Results

In order to test the hypotheses of the study, a number of different statistical analyses were carried out. Task structure, pre-task planning time and proficiency level were the independent variables, while 12 measures of fluency, accuracy and complexity were the independent variables. Factor analyses were run to investigate whether the dependent variables truly represented distinct factors. Based on the results of the factor analysis, a repeated measures MANOVA was performed to test the overall effect of the independent variables on the dependent variables. Finally a series of ANOVAs and t-tests were run to exam-

ine the differences among tasks, between the planning conditions and language proficiency levels.

Underlying factors in language performance

Separate factor analyses were run for each of the four tasks with the 12 measures of fluency, accuracy and complexity. Prior to the analysis, the suitability of the data for factor analysis was investigated. Inspection of the correlation matrixes revealed the presence of many coefficients of .4 and above. The Kaiser-Meyer-Olkin values were above .68 for all the tasks, exceeding the recommended value of .60 (Kaiser 1974). Barlett's Test of Sphericity reached statistical significance, supporting the factorability of each correlation matrix. The factor structure obtained was remarkably similar for all four tasks, and so only the results for the Football task only are shown here, as Table 4, to save space.

As Table 4 shows, factor 1 is made up of six high loadings on measures of length of run, speech rate, total amount of silence, total time spent speaking, number of pauses and length of pauses. These measures refer to different temporal aspects of fluency and suggest one relatively unified dimension. This hypothetically means that the more fluent participants would be expected to use a significantly greater length of run, a faster speech rate, less silence, fewer pauses, shorter pauses as well as more time spent speaking in their performance. The result of the factor analysis in Mehnert's study (1998) supports the same loadings for speech rate, length of run and total amount of silence.

The second factor loading is based on reformulations, false starts, replacements and repetitions, measures associated with another aspect of fluency, i.e. repair fluency (Freed 2000; Skehan 2001, 2003). The loadings for reformulation and false starts define the factor most clearly and replacement and repetition follow with lower, yet significant loadings. (This pattern was identical across all four tasks.)

The results of the factor analyses indicate that accuracy and complexity, together with length of run, load highly on the third factor suggesting that more accurate language was also more complex. Furthermore, the fact that these measures are associated with each other indicates that they are reflecting the same underlying constructs. This confirms the assumption that accuracy and complexity are both aspects of form and are in contrast with fluency as an aspect of meaning. Length of run loaded on this factor for the Football and Picnic tasks, but not in the factor analysis for the two less structured tasks.

Table 4. Factor analysis for the football task

Measures	Factor 1	Factor 2	Factor 3	Communality
Reformulations		.88		.880
False starts		.94		.892
Replacements		.41		.276
Repetitions		.62		.490
Accuracy			.65	.662
Complexity			.87	.716
Length of run	-.66	-.44	.43	.767
Speech rate	-.84			.793
Total silence	.95			.912
Time spent speaking	-.94			.902
No. of pauses	.80			.736
Mean length of pause	.87			.844

The effects of task structure, planning condition and proficiency level on language performance

To investigate the experimental effects a repeated measures MANOVA was carried out. Based on the results of the factor analyses, four measures were selected from the total number of 12 measures to represent the dependent variables: number of false starts, number of pauses, accuracy and complexity of the performance. The criterion for selecting one measure from the temporal and one from the repair fluency in each factor group was the consistency in loadings of these measures across all the tasks. As regards language form, since complexity and accuracy have shown themselves in previous research to be influenced by different independent variables, they were both included in the analysis. The independent variables of the analysis were planning and proficiency level, each with two levels and task structure with four levels. The results from the repeated measures MANOVA are presented in Table 5.

With regard to the between-participants effect, the analysis revealed a significant difference between the planners and non-planners (Pillais = .179, $F = 4.00$, $P = .006$) as well as low and high proficiency levels (Pillais = .374, $F = 10.89$, $P = .001$). A significant difference was further seen across the four tasks for the within-participants variable (Pillais = .754, $F = 16.78$, $P = .001$) with differences being concentrated in the number of pauses (Walkman: $M = 26$, Unlucky Man: $M = 22$, Picnic: $M = 19$, Football: $M = 18$); in complexity (Walkman: $M = 1.36$, Unlucky Man: $M = 1.32$, Picnic: $M = 1.60$, Football: $M = 1.43$); in false starts (Walkman: $M = 5.2$, Unlucky Man: $M = 4.41$, Picnic: $M = 4.59$, Football: $M = 3.97$); and in accuracy (Walkman: $M = .30$, Unlucky Man:

Table 5. Results of repeated measures MANOVA

Effects	Pillai's Value	F	BGdf	WGdf	Sig.
<i>Between-participants effect</i>					
Planning	.179	4.00	4	73	.006*
Proficiency	.374	10.89	4	73	.001*
Planning × Proficiency	.103	2.09	4	73	.09
<i>Within-participants effects</i>					
Task	.754	16.78	12	65	.001*
Task × Planning	.263	1.93	12	65	.16
Task × Proficiency	.278	2.09	12	65	.12
Task × Planning x proficiency	.288	2.19	12	65	.09

*Significant difference is reached

Table 6. Univariate test of within-participant effect

Source	Measure	Sum of squares	df	Mean square	F	Sig.
Task	No. of pauses	3047.55	3	1015.85	20.21	.001*
	Complexity	3.53	3	1.18	25.65	.001*
	False start	63.00	3	21.00	3.95	.009*
	Accuracy	1.22	3	.407	29.80	.001*
Task × Planning	No. of pauses	64.55	3	21.52	.42	.73
	Complexity	.431	3	.144	3.126	.02
	False start	24.85	3	8.28	1.55	.2
	Accuracy	.082	3	.027	2.015	.11
Task × proficiency	No. of pauses	6.83	3	2.27	.045	.98
	Complexity	.465	3	.155	3.37	.01*
	False start	21.60	3	7.20	1.35	.25
	Accuracy	.123	3	.041	3.00	.03
Task × Pl. × Prof.	No. of pauses	291.9	3	97.3	1.93	.124
	Complexity	.184	3	.06	1.33	.26
	False start	29.18	3	9.72	1.83	.14
	Accuracy	.034	3	.011	.83	.36

M = .30, Picnic: M = .43, Football: M = .42). When the results for the dependent variables were considered separately through a Univariate F test, using a Bonferroni adjusted alpha level (recommended by Tabachnic & Fidell 1996), significance was reached for all the four measures as a result of the task effect. However, the only significant result in the interaction effects between task and

proficiency level was seen for complexity. Results of the Univariate F test are provided in Table 6.

The results indicate that there is a statistically significant difference across the tasks. Further comparisons of all the four measures showed that the structured tasks were not different from one another but they were significantly different from the unstructured tasks in terms of the number of pauses and accuracy. Regarding complexity and false starts, the unstructured tasks were not significantly different from one another but were different from one of the structured tasks.

The effects of task structure

A one-way ANOVA was carried out on each of the 12 independent variables to determine which measures yielded significant differences. Where significance was reached the Scheffé test was run to establish where the differences were located. In cases of non-significant results, pairwise comparisons between tasks were run to explore the differences between pairs of the tasks to gain an understanding of trends within the data. Results of the ANOVAs for all the tasks are given in Table 7 with the F-values, significance levels, means for the four tasks, standard deviations, and an indication of where differences reached significance.

The results show that differences across the four tasks were significant on the measures of total amount of silence, length of run, speaking time, number of pauses and false start (See Table 7). For all these measures the differences reached significance with performance on one or both of the structured tasks being more fluent than performance on one or both of the unstructured tasks. For number of pauses and speaking time the two structured tasks, i.e. Football and Picnic, were significantly more fluent than the two unstructured tasks, i.e. Unlucky Man and Walkman. For length of run Football was significantly more fluent than Unlucky Man. For total amount of silence Football and Picnic are different from Walkman and Unlucky Man, and for false starts Football was significantly more fluent than Walkman. The results of the ANOVAs reveal that significant differences were reached between the structured tasks and unstructured tasks with regard to the accuracy measure ($F = 9.79$, $P < .001$). The results of the Scheffé test showed that the two structured tasks generated significantly more accurate language than the two unstructured tasks. Although it was hypothesized that there would be no significant difference between the complexity of the language generated by the structured tasks and that of the un-

Table 7. Results from the ANOVA on all measures for the four tasks

Measures	F	P	Task			Structure		Location of Sig. differences	Sig. Pairwise Comparison
			Walkman	Unlucky Man	Picnic	Football			
Total silence	3.80	.04*	29.62 (SD = 23.87)	27.39 (SD = 25.09)	20.57 (SD = 22.72)	19.47 (SD = 19.56)	F vs. W	F P vs. U W	
Length of run	4.99	.008*	3.59 (SD = 1.07)	3.29 (SD = 1.09)	3.85 (SD = 1.41)	4.05 (SD = 1.57)	F P vs. U	F vs. U W	
Pause length	2.72	.16	1.02 (SD = .54)	1.09 (SD = .61)	.9 (SD = .38)	.9 (SD = .4)		F vs. U	
No of pauses	6.90	.001*	26.6 (SD = 12.71)	22.51 (SD = 11.46)	19.92 (SD = 12.41)	18.47 (SD = 11.89)	F P vs. W	F P vs. U W	
Speaking time	6.45	.001*	.71 (SD = .15)	.69 (SD = .15)	.76 (SD = .13)	.79 (SD = .13)	F P vs. U	F P vs. U W	
Speech rate	1.57	.19	94.66 (SD = 29.42)	87.76 (SD = 29.95)	99.27 (SD = 41.77)	94.85 (SD = 33.09)			
False start	1.87	.13	5.2 (SD = 3.88)	4.41 (SD = 3.35)	4.58 (SD = 3.05)	3.96 (SD = 3.02)		F vs. W	
Reformulation		2.79	3.06 (SD = 2.4)	2.28 (SD = 1.96)	2.79 (SD = 2.09)	2.21 (SD = 2.17)		F vs. W U vs. W	
Replacement	1.29	.27	.61 (SD = .83)	.71 (SD = 1.41)	.43 (SD = .69)	.48 (SD = .79)			
Repetition	.53	.65	4.28 (SD = 4.58)	3.88 (SD = 4.69)	3.46 (SD = 3.45)	3.78 (SD = 3.70)			
Accuracy	9.79	.001*	.30 (SD = .20)	.30 (SD = .18)	.43 (SD = .19)	.42 (SD = .22)	F P vs. U W		
Complexity	15.19	.001*	1.36 (SD = .28)	1.31 (SD = .20)	1.59 (SD = .33)	1.43 (SD = .28)	P vs. F U W	P vs. U W	

* Significant differences are reached across tasks; F = Football, P = Picnic, U = Unlucky Man, W = Walkman.

structured tasks, the results indicate that the Picnic task ($M = 1.59$) generated significantly greater complexity of language compared with the other tasks.

The effects of strategic planning and proficiency

Hypothesis Three predicted that language performance under planned conditions would be more fluent, more accurate and more complex than that produced under unplanned conditions. A series of *t*-tests were carried out on each dependent variable to locate the effect of strategic planning time on different measures of fluency, accuracy and complexity. Furthermore, to compare the effect of planning with the effect of language proficiency on the dependent variables, a number of *t*-tests were carried out on all measures of fluency, accuracy and complexity for the two levels of proficiency. The results of the *t*-tests for planning conditions and proficiency levels are presented in Table 8.

The results of the *t*-tests show that the effect of strategic planning reached statistical significance for measures of total silence ($t = 4.16, P = .001$), length of run ($t = 4.16, P = .001$), pause length ($t = 5.93, P = .001$), speaking time ($t = 5.80, P = .001$) and speech rate ($t = 3.14, P = .008$). The mean scores for each measure show that performances were significantly more fluent under planned conditions. Although the measures of number of pauses and reformulations do not reach statistical significance, the reductions in these measures clearly show that performance under planned conditions tends to be more fluent than unplanned performance.

All measures of temporal fluency are significantly higher in the intermediate language proficiency group indicating that the language performance of high proficiency participants is more fluent than the performance of low proficiency participants. Interestingly, the effect of strategic planning on the total amount of silence, pause length and speaking time is greater than the effect of language proficiency, i.e. it appears to be better to be a low proficiency planner than an intermediate proficiency non-planner.

It was further hypothesized that language performance would be more accurate under planned conditions. Results of the *t*-tests show that accuracy significantly improved under the planned condition ($t = 5.52, P = .001$). Language performed by high proficiency participants is also significantly more accurate than low proficiency participants' language ($t = 7.34, P = .001$). However, the effect of proficiency level on accuracy is greater than the effect of pre-task planning.

Language produced under planned conditions was predicted to be more complex under planned conditions. As can be seen in Table 8, a statistically

Table 8. Results of T-test for planning conditions and proficiency levels

Measures	T	P	Unplanned	Planned	T	P	Elementary	Intermediate
Total silence	4.16	.001*	29.53 (SD = 27.42)	19 (SD = 16.55)	3.07	.004*	28.22 (SD = 22.94)	20.32 (SD = 22.88)
Length of run	4.16	.001*	3.39 (SD = 1.18)	4.00 (SD = 1.4)	6.12	.001*	3.26 (SD = 1.10)	4.12 (SD = 1.46)
Pause length	5.93	.001*	1.14 (SD = .6)	.81 (SD = .3)	3.72	.001*	1.08 (SD = .56)	.87 (SD = .45)
No. of pauses	1.68	.18	23.05 (SD = 12.91)	20.70 (SD = 11.92)	2.00	.08*	23.26 (SD = 11.28)	20.48 (SD = 13.42)
Speaking time	5.80	.001*	.69 (SD = .15)	.78 (SD = .13)	2.65	.01*	.71 (SD = .15)	.76 (SD = .14)
Speech rate	3.14	.008*	88.23 (SD = 38.04)	100.04 (SD = 28.36)	6.72	.001*	82.15 (SD = 25.82)	106.12 (SD = 36.94)
False start	.21	.82	4.5 (SD = 3.17)	4.58 (SD = 3.55)	2.87	.008*	5.07 (SD = 3.63)	4.00 (SD = 3.00)
Reformulation	1.12	.26	2.45 (SD = 2.03)	2.72 (SD = 2.33)	2.46	.03*	2.88 (SD = 2.53)	2.28 (SD = 1.74)
Replacement	.57	.56	.53 (SD = 1.12)	.59 (SD = .81)	1.95	.1	.66 (SD = 1.18)	.45 (SD = .70)
Repetition	.78	.43	3.67 (SD = 3.57)	4.00 (SD = 4.63)	1.35	.34	4.16 (SD = 3.82)	3.54 (SD = 4.41)
Accuracy	5.52	.001*	.30 (SD = .19)	.42 (SD = .21)	7.43	.001*	.28 (SD = .19)	.44 (SD = .20)
Complexity	2.23	.04*	1.38 (SD = .29)	1.46 (SD = .29)	6.62	.001*	1.32 (SD = .23)	1.53 (SD = .31)

* Significant differences are reached.

significant difference is reached for the complexity of performance between the two planning conditions ($t = 2.32$, $P = .04$) with the planned group achieving higher degrees of complexity in their performance. The results of the t -tests also reveal that the effect of proficiency level on complexity seems to be greater than the effect of pre-task planning ($t = 6.62$, $P = .001$).

Perceptions of task difficulty

A three-way ANOVA, using responses to the task difficulty questionnaire items as the dependent variable and task structure, pre-task planning and proficiency levels as the independent variables, was carried out. Considering the Bonferroni adjusted alpha level (Tabachnic & Fidell 1996), the results of the three-way ANOVA show a significant difference for task structure ($F = 32.63$, $P = .001$) and also a significant difference for the planning conditions ($F = 6.11$, $P = .02$). However, no significance was reached for proficiency level or the interaction between the dependent variables. Table 9 shows the results of the three-way ANOVA on participant perceptions of task difficulty.

Mean scores of the perceptions of difficulty across tasks under the two pre-task planning conditions are shown in Table 10. The comparison shows that the two unstructured tasks, i.e. the Unlucky Man and Walkman tasks, were rated as more difficult than the two structured tasks under both the planning conditions.

Table 9. Three-way ANOVA on perceptions of task difficulty

Source	Type III Sum of Squares	df	Mean Square	<i>F</i>	<i>P</i>	Eta. Square
Task	42.10	3	14.03	32.63	.001	.244
Planning	2.62	1	2.62	6.11	.02	.02
Proficiency Level	1.12	1	1.12	2.62	.106	.009
Planning \times Task	.58	3	.19	.45	.71	.004
Task \times Prof.	.93	3	.311	.72	.53	.007
Prof. \times Planning	.37	1	.37	.87	.34	.003
Plan \times Prof. \times Task	.33	3	.11	.25	.85	.003

Table 10. Mean scores of perceptions on task difficulty

Tasks	Football	Picnic	Unlucky Man	Walkman
Unplanned	1.90	1.95	2.67	2.55
Planned	1.80	1.62	2.52	2.40

Table 11. Three-way ANOVA on usefulness of planning time

Source	Type III Sum of Squares	df	Mean Square	<i>F</i>	<i>P</i>	Eta. Square
Task	1.53	3	.51	.81	.48	.016
Proficiency	.94	1	.94	1.50	.22	.010
Task × Prof. Level	6.25	3	2.08	3.33	.06	.063

A Scheffé test was then carried out to explore where the significant differences were located across the tasks. The multiple comparisons showed that the two structured tasks, i.e. the Football and Picnic tasks were not statistically different from one another but were statistically different from the two unstructured tasks.

Questionnaires of the planned group included a section on the usefulness of the strategic planning time for each of the tasks. A two-way ANOVA was carried out to investigate whether the participants from the two proficiency levels found pre-task planning time more useful for one or some of the tasks. Results of the analysis did not reveal any significant differences across the tasks or the proficiency levels. Table 11 shows the results of the analysis.

Discussion

At the outset, the results of the factor analyses are worth brief comment. First, the consistency of the results across the four tasks is noteworthy, and provides some confidence in the robustness of the structures that are involved. Second, the results regarding fluency are striking, in that there is a separation between repair fluency, on the one hand, and breakdown fluency/rate of speech/unit size, on the other. The repair fluency group loads together consistently, and separately from other aspects of fluency, suggesting that a concern to modify utterances on-line is somewhat distinct from a capacity to organise speech in real time. In addition, the grouping together of the three non-repair aspects of fluency (avoidance of pausing; speed of production; and size of unit that is produced) suggests that while these areas may be distinguished conceptually, in actual performance they draw upon one general capacity to orchestrate speech effectively. Third, it is interesting that the two form-focussed areas of accuracy and complexity load together, consistently as the third factor, suggesting that a concern for language form is less prominent in the data matrix, and does not differentiate so clearly between these two different aspects of performance as

is the case in other task-based studies, e.g. Skehan and Foster (1997). We will return to the relationship between these two below.

The findings regarding task are particularly interesting. The figures for silence and pausing are clear, since they each reduce (i.e. indicate greater fluency) as a function of hypothesised task structure. For example, number of pauses reduces from 26.6 (Walkman) to 22.51 (Unlucky Man) to 19.92 (Picnic) to 18.47 (Football). False starts and reformulations pattern similarly, and it could be argued that the trends are similar for replacement and repetition. It is interesting that the other aspects of the bundle of measures which defined the first factor do not pattern so clearly, and it appears that the putatively more structured Picnic and Football tasks lead to somewhat higher performances than do the less structured tasks, especially Unlucky Man, for length of run, speaking time, and speech rate. It appears that fluency is a complex construct and that explaining these patterns will be an interesting challenge for future research and theorising.

The results for accuracy suggest that the two least structured tasks, (Walkman and Unlucky Man) contrast with the two more structured tasks (Picnic and Football) but that within each pair of tasks, there is little difference. In a sense, therefore, this does provide support for the claim that task structure influences accuracy, but there is not the same general progression over all four tasks as there is with breakdown fluency. In other words, there is not evidence in support of a scale of structure, with gradations. The evidence is more consistent with a threshold of structure, such that above this threshold, accuracy is supported, but below it, accuracy is reduced. If the Football task contains additional structure to the Picnic task, this does not manifest itself in more accurate performance. It may be necessary therefore to rethink how structure impacts on the accuracy of performance in greater detail.

Regarding complexity, there are further complications in the results. The prediction was that there would be no effects for complexity. In the event, there is a significant finding, in that the Picnic task has generated greater complexity than the other three tasks. None of the other tasks differ from one another. This result presents a puzzle, and will have to await further research. However, it could be argued that a post-hoc analysis of the different picture series does suggest that an additional feature distinguishes the Picnic task from the other three. This is that the point of this story sequence requires subjects to make connections between foreground and background elements in each of the various pictures in the sequence, and that it is this which underlies the greater complexity of the language used. Future research will be needed to explicate this finding.

The findings on planning are fairly consistent and clear. Fluency is significantly advantaged by strategic planning. Regarding breakdown fluency, this applies to total silence, and pause length, but not to number of pauses. It applies consistently to rate of speech and size of unit indices (speaking time, speech rate, and length of run). None of the repair fluency measures produces significance, and interestingly each is higher for the *planned* condition, i.e. the planners seem to be more likely to engage in modification of speech on-line. But taken broadly, these results indicate strong support for the claim that pre-task planning leads to a significant increase in the main dimensions of fluency.

Turning to the form-linked measures, the results for complexity are significant, though somewhat surprisingly only at the 0.05 level, with a difference of 1.46 to 1.38, measured in AS-units. This does suggest that pre-task planning produces greater complexity, but the size of the effect is not especially great. On the other hand, the results for accuracy reach a much more demanding significance level, and indicate that pre-task planning time has much more of an impact on accuracy. Bearing in mind that there is an assessment context for these results, it may be the case the learners taking tests shift priorities somewhat, and value accuracy more than complexity, whereas learners doing tasks as part of classroom behaviour may be more inclined to focus attention on doing a task to its potential, and allocate attention to complexity.

The findings for proficiency level are mostly straightforward, and provide interesting validity confirmation that this data elicitation format produces results consistent with conventional test results. The repair fluency measures present their usual mixed picture, in this case with false starts and reformulations being significantly less frequent in the higher proficiency group, and with no significances for replacement and repetition. Otherwise, all measures except number of pauses produce significant results in favour of the higher proficiency group, with all these significances being at least at the 0.01 level. The effect sizes for accuracy and complexity are particularly noteworthy.

There is one particular interesting finding when one looks at the relationship between pre-task planning and proficiency level. As indicated earlier, there are occasions where there is higher performance by the Elementary proficiency planners compared to the Intermediate non-planners. This is interesting because it suggests that higher performance can be achieved if task/assessment conditions allow for planning compared to simply having a higher proficiency level. If we relate this to the model of spoken language assessment shown in Figure One, it makes it clear that scores assigned may not reflect simply profi-

ciency level, but also the conditions under which a task is done. We will return to this below.

Finally, the results for perceptions of task difficulty are worth discussing. Broadly, the unstructured tasks (Unlucky Man and Walkman) were rated as more difficult than the structured tasks (Football and Picnic), with there being little difference between the pairs of tasks in each case. Further, the non-planners rated the task as more difficult than did the planners, although this difference was not so great in scale. In other words, participants rated as *less* difficult the tasks and task conditions when they did better. If we assume that they did better because they were more able to do justice to what they perceived as their “true” proficiency, then it would appear that the participants themselves were aware, at some level, that they were performing differently, and presumably, in a more satisfactory manner. It would appear, in other words, that having either a clear general structure within which to frame the narrative, or having pre-task planning opportunity to enable resources to be mobilised (and possibly achieve the same end) led to learners who felt they were more in control of the situation (cf. awareness of Formulator operation). It is particularly interesting that the lowest difficulty rating, i.e. the “easiest” task, was the Picnic story under the planned condition. This is interesting, and worthy of future research, because this was the experimental condition which led not simply to higher fluency and accuracy, but also complexity. Despite interpreting the task in such a way that the most advanced language was produced, participants nonetheless regarded the task as easier.

Conclusions

Above all, the present study has contributed results which clarify the functioning of the model shown in Figure 1, and help to take the model beyond schematic value and towards a sounder empirical base. The study clarifies that:

- task structure is an important influence on performance;
- strategic planning has an effective and predictable influence, and generally improves the level of performance which results. This applies to all aspects of performance, including accuracy;
- different aspects of performance are affected slightly differently by structure and pre-task strategic planning.

These results suggest that it is fruitful to chart potential task-based influences on actual candidate performance. The wider issue, clearly, is that tasks and

task conditions vary, with the result that a particular testing encounter could use a combination of circumstances which inadvertently impact upon performance, and subsequent assessments. Without knowledge of these effects, the danger is that test scores which are assigned are partly artifactual, and difficult to compare with results obtained under different conditions.

The results also indicate how more experimental studies can contribute to language test validation. It may be the case that some variables, e.g. adequacy, perspective, immediacy, as defined by Iwashita et al. (2001) do not have significant effects. The present study has shown that there are, though, other relevant variables (task structure) which do impact upon performance. Clearly, further research to uncover other potential influences is warranted. But the present results also question the findings regarding the non-influence of strategic planning on accuracy which have occasionally been reported in the literature (e.g. Crookes 1989; Ortega 1999). In the present case, strategic planning generated consistent significances, even, on occasion, leading to stronger influence on performance than proficiency. Future research will have to explore why strategic planning works in some circumstances and not in others.

The present results are also compatible with the revised system proposed by Brown et al. (2002) for assessing task difficulty. Amongst the factors that they propose that can influence the difficulty of tasks they include the cognitive operations that are required to carry out a task, and within this area, they have sub-headings of input/output organisation and input/output availability. The first concerns the degree of transformation of the elements of the task that are required, and the second focuses on the information that is the basis for the task and the ease of accessibility of this information. The findings on structure are relevant to each of these. More clearly structured tasks contain a clearer organisational framework, removing the need for attention to be directed to re-organising material. As a consequence, attention becomes available to access information which will enable the task to be carried out more easily. In other words, greater structure does appear to ease the task faced by the test-taker, and permits the allocation of attention to formulation (Levelt 1989), and as a result, fluency and accuracy. The findings, in other words, suggest slight modification of these two sub-areas of cognitive operations in the system provided by Brown et al. (2002), so that within their sub-category of input/output organisation, information structure is highlighted a little more than it is at present.

Two additional problems, though, are worth mentioning. First, there is the issue of task difficulty itself. Brown et al. (2002) regard the problem of identifying task difficulty to be a central aim for research. They regard difficulty to be a joint function of ability requirements and task characteristics. In other words

they want to ask a question like: “What difficulty level in a task can a candidate of a certain proficiency level handle?” In this way, in a testing situation, one would want to give candidates of different proficiency levels tasks of appropriate difficulty level to transact. The aim of testing would then be to most efficiently identify the maximum level a particular test candidate could cope with. But the present research has once again confirmed that performance is multi-dimensional, (cf. the degree of independence between performance areas such as fluency, accuracy, and complexity) and compensatory and that it is difficult to propose that there is a central “gold standard” criterion one can use to identify difficulty level. Task characteristics may vary and this variation may connect systematically with different aspects of performance, but the problem is that these different dimensions do not function in unison: increasing performance in one area may not be associated with increased performance in another. The problem then becomes how to handle this inconsistency at the level of the ratings and measurements that are at play. Perhaps this is what Bachman (2002) means when he argues that difficulty is not a separate factor, but resides in the interactions of all the features that are involved in an assessment situation.

The second general problem may be a version of the Observer’s Paradox. It is clear that there are inconsistencies in the findings between different studies in the literature and that one of the goals of future research will be to account for such discrepancies. In that respect, one can distinguish between two sorts of research:

- studies which are pedagogy or acquisition oriented;
- studies which replicate testing conditions.

The inconsistency between these two areas in terms of the likelihood of finding significant effects for experimental variables is worth pursuing. Potentially, four sources of difference may be relevant:

that testing-linked task research leads to a different type of language use, because of the prominence of the assessment context, and possibly a greater focus on accuracy (since this is what tests are traditionally perceived to be about). This might mean that there is less scope for experimental manipulations to produce additional accuracy effects, since attention is already being directed to that area of performance. Even so, this would not explain why studies such as Iwashita et al. (2001) have failed to find complexity or fluency effects either. However, a broader Observer’s Paradox interpretation would presumably be arguing for

inconsistency going beyond accuracy, and incorporating other domains as well. That the act of testing distorts performance remains a worrying possibility.

that different types of experimental variables are researched in the two contexts, and that it is the variables associated with acquisition or pedagogy (A/P) which are more likely to produce significant effects. Certainly the A/P studies have used a greater range of tasks, including decision-making, narrative, personal information exchange, picture description and so on, compared to the more narrative-task focus in assessment. This has permitted the use of variables which emphasise the interactive nature of the wider range of tasks. Conversely, assessment oriented researchers have tended to choose variables which can be manipulated to generate potential differences in difficulty level, and act as a blueprint for generating multiple versions of tasks of equivalent difficulty level in a straightforward manner. Hence the use of variables such as adequacy and perspective. But these differences in findings do not account for the discrepancies in the effects of planning in the two areas, or in variables such as immediacy of information, which have yielded significant findings for Foster and Skehan (1996) and Robinson (2001c) but not for Iwashita et al. (2001). There seem to be other, unexplored variables at play which need to be probed further.

that common variables have been operationalised differently in the two contexts, with the result that inappropriate comparisons are being made. There is certainly some evidence that this is the case. Wigglesworth (2001), for example, operationalises structure in terms of the amount of information which is provided, an approach not dissimilar to Iwashita et al.'s (2001) treatment of *adequacy*. Both approaches certainly contrast with Skehan and Foster's (1999) treatment of structure in terms of information organisation, as well as the approach taken in this chapter. But there are other variables which are operationalised similarly in both research contexts, such as Iwashita et al.'s (2001) use of immediacy, compared to Robinson's (2001c) "here and now" vs. "there and then" conditions, or alternatively, the approaches to planning in the various studies, and here the testing studies, with lack of significance, compare markedly with the non-assessment studies which do report significances. Once again we are left with a puzzle.

that scoring procedures differ in the two contexts, with A/L researchers tending to use detailed measures and assessment researchers using ratings. While it is certainly true that A/L researchers do not use ratings of performances (although it would not be a bad thing if they did), the assessment researchers have fre-

quently used a range of measures and the lack of significances are reported with *both* methods of evaluating performance.

Perhaps the conclusion to draw here is the familiar one that more research is needed. Although it would be worrying to conclude that it is the Observer's Paradox that strikes at the heart of testing, so to speak, it may be premature to come to that conclusion yet. It appears that the discrepancies in the results, while unlikely to be the result of the differences in scoring procedure, could well be produced by the different foci and motivations for experimental variables, as well as the different operationalisations that have been used. The present study does provide some very relevant results, indicating clear differences between conditions, and appreciable effect sizes, and these in a context which was approached as an assessment environment. Perhaps the best recommendation would be that, with research in this area still developing techniques and understandings, the role of A/L work will be to offer suggestions for relevant variables and the standardization of operationalisations of variables, but that these insights and findings will need to be confirmed within assessment contexts. It is to be hoped that the approach to structure and pre-task planning portrayed in the present chapter can make a contribution to this.

Notes

* The authors would like to thank Rod Ellis and two anonymous reviewers for reading earlier drafts of this chapter, and offering comments which have enabled us to strengthen it considerably.

1. Even so, it should be noted that Robinson might well have predicted this result, since he argues (see above), that language complexity is driven by functional complexity.

SECTION VI

Conclusion

In the main, the view of planning evidenced in the preceding chapters has been essentially cognitive. That is, authors have viewed planning as a cognitive process that will have cognitive implications for the way in which learners subsequently perform the task. With a few exceptions (Ortega's and Sanguran's chapters, in particular) the actual cognitive processes involved in both planning and in task performance have not been directly investigated but rather have been inferred from analysis of the texts produced by the learners. Further, the method of analysis strongly preferred by these authors has been linguistic (i.e. measures of fluency, accuracy and complexity). It is fitting, therefore, that the concluding chapter to this book should point out the limitations of this paradigm and emphasise the need for a broader perspective, one that takes into account the fact that both planning and task performance are social as well as cognitive activities.

At the heart of Batstone's socio-cognitive view of tasks is the idea that learners can orientate to both task planning and task performance in one or two ways – either in terms of communicative discourse (where the emphasis is accomplishing the task with economy or efficiency) or as learner discourse (where the emphasis is on taking risks for the purpose of interlanguage development). Batstone opposes the standard view in SLA research, namely that interlanguage development is best served when learners engage in communicative discourse that causes them to link form to meaning. He argues that, in fact, it may be better served when they are released from the need to adopt a purely communicative approach and instead see the task (and planning for the task) as an opportunity to stretch their linguistic resources by being adventurous.

In this respect, Batstone makes an interesting point about the design of narrative tasks – the task type that has predominated in the studies reported in this book. He points out that such tasks can be designed to be performed with an interlocutor (as in Ortega's task in Chapter 3) or without an interlocutor (as in Ellis and Yuan's task in Chapter 6). The justification for including an interlocutor is that it makes the task more communicative. Batstone suggests, however, that there may be an advantage in excluding an interlocutor, as this

may predispose learners to engage in learner rather than communicative discourse. A study that compared learners' performance in a narrative task with and without an interlocutor may be an excellent way of investigating Batstone's central thesis – that the type of discourse that arises will depend crucially on the learners' socio-cognitive orientation.

CHAPTER 10

Planning as discourse activity

A sociocognitive view

Rob Batstone

University of Auckland

Planning and language learning: Automaticity

SLA research into planning over the past fifteen years has established itself as a line of enquiry with very considerable promise, not least because it suggests how a relatively uncomplicated procedure in terms of pedagogy can have considerable and varied payoffs in terms of language learning. In very broad terms, the picture which is emerging suggests that through the use of carefully designed tasks, we can considerably expand learners' attentional capacity, enabling them to engage selectively with language in ways which lead them into being more accurate, more complex and/or more fluent in their performance (see the summary in Ellis 2003: 127–136 and Chapter 1). More speculatively, but also more significantly, it is hoped that these short-term gains in language production will contribute in the longer term to tangible advances in learning.

One such advance concerns the cumulative effect of learners exploiting planning (both strategic and on-line) to use language which is relatively complex and at the 'cutting edge' of their current interlanguage capabilities. Choosing to use output which is complex relative to the learners' current capacity, an ability sometimes referred to as 'pushed output' (Swain 1995), is said to enable learners gradually to automatize their knowledge and hence call upon this language with greater control and facility during language use (Bialystok 1990; Anderson 2000). It is the capacity of planning to influence this particular aspect of language learning which is the main focus of the present article.

Towards a discourse/sociocognitive reconceptualization of planning

Planning, as it is typically researched, gives us little if any insight into the discourse motivations of the learners it involves. Rather than adopt a discourse-sensitive approach, SLA researchers have preferred to take a more cognitive stance, to represent learners largely as information processors, and to conceive of planning essentially as a means of freeing up attentional resources so as to enable learners to pay more attention to form (Skehan 1998b). The rest of this chapter is devoted to providing a tentative critique of this perspective.

In line with recent work in sociolinguistics generally (and from sociocultural theorists in particular), it is argued that cognitive perspectives fail to give due credit to such matters as the importance of social context or learner identity (e.g. Firth & Wagner 1997; Rampton 1997; Lantolf 2000; Block 2003).

But this does not mean that we should simply abandon the gains made through cognitive theory. The argument that planning has the potential to enhance learners' attentional capacity, for example, is a powerful one which is grounded in a great deal of SLA research (in particular Schmidt 1990; 2001). But there are limits to what we can usefully say about attention if we leave it in a social vacuum, and there are good grounds for integrating attention within a wider, discourse perspective.

How might such an integration be realized with respect to second language planning? Cognitive theorists have argued very clearly for linkage between planning and a corresponding reduction in attentional capacity. What is less commonly argued is the notion that the effects of planning on attention are as much a matter of social action as they are of cognitive processing. As will be argued shortly, both the learners' capacity to plan effectively and their ability to act on planning by 'pushing' their output are rooted in social context. The former presupposes some prior experience of commensurate educational culture and related classroom discourses, whilst the latter presumes a capacity to act on and within discourse in socially assertive and potentially face threatening ways. Thus we can usefully think of learners' engagements with language through planning as being *sociocognitive*: 'cognitive' because attention is so centrally implicated, 'social' because attention is activated through discourse endeavour of very particular kinds, 'sociocognitive' because the cognitive and the social are so closely intertwined.

So what kind of integration between the social and the cognitive might be most productive for second language planning? What is needed, perhaps, is an approach to planning where social factors (such as the learners preparedness to take linguistic risks and be assertive in their discourse) are *congruent* with cog-

nitive factors (in particular the need to enhance attentional capacity). Much of the existing research on planning assumes that this kind of congruence is best served through communicative discourse where the primary goal is the signalling of necessary meanings which are transactionally warranted. But it is suggested here that learners may be better served when they interpret planning in terms of a 'learning discourse'. In a learning discourse the primary goal is explicitly to push one's output for a learning purpose, and therefore to go beyond communicative boundaries on what it is appropriate or relevant to say. Viewed sociocognitively, a learning discourse calls for the creation of a social context which supports risk taking and which rewards output pushing. Data from a number of planning studies are re-examined in order to provide tentative evidence for the efficacy of a 'learning' orientation to discourse relative to its communicative counterpart.

Evidence of pushed output? Issues around the definition of 'complexity'

But we begin with a necessary preliminary. If planning is to lead to payoffs in terms of learning, learners need (amongst other things) to push their output by using challenging, relatively complex language. Indeed, it is on account of this kind of linguistic risk taking that social context looms so large. But most of the extant research on planning does not convincingly demonstrate that planning necessarily leads to output pushing. To date, scholars have relied almost entirely on the presentation of abstracted quantitative data as evidence for the impact of planning on learners' language production. But this kind of data fails fully to demonstrate that learners are in fact 'pushing' their language to the extent that they could or (perhaps) that they should be doing in order to maximize their learning.

Based on the work of Skehan (1996a, 1998b), Foster and Skehan (1999: 211) note that complexity reflects "a greater willingness to take risks and to try out more exploratory language", and is thus closely related to pushed output. In linguistic terms, complexity has been operationalized using a variety of measures, ranging from the relatively narrow (e.g. the use of subordination: Wigglesworth 1997) through to the relatively broad (e.g. the use of both T-units and C-units: Mehnert 1998).

But it is not clear how sensitive these systems would be in the case, say, of the learner who would need to push her cognitive resources to the limit just to grammaticize indirect object constructions, or word order distinctions between subject and object. For such a learner, these forms may well mark the

boundary of her current competence and hence be relatively complex (i.e. complex relative to her current ability). But in empirical terms this learner's level of complexity is either low or zero (depending on the exact scheme being used). In short, there is a confusion here between linguistic tokens of complexity (such as subordination, as determined through standardized empirical measures), and *psycholinguistic* tokens (i.e. language related to its cognitive degree of difficulty for the individual learner). The two need to be kept distinct, but in much SLA planning research there is a tendency to equate them, and to treat linguistic measures of complexity as if they were psycholinguistically valid.

There are similar issues with learners operating at a higher level of proficiency. Take, for example, a learner who uses a relatively large amount of complex dependent clause structure as a result of prior strategic planning, and in comparison to the 'no planning' condition (i.e. a task where the learner was given no opportunity to plan). It may well be the case that some of this language is 'cutting edge' and so constitutes genuine pushed output. But not all, because a lot of this learner's 'complex' output is already well on the way to automaticity and is accessed with some ease. Yet it will all get marked up the same way in the researcher's data: as (linguistically) complex, and so as evidence of the 'pushing power' of the task in question. Here again there is a confusion between linguistic, generalized benchmarks of complexity on the one hand, and psycholinguistic, learner-dependent measures on the other.

The use of generalized linguistic measures of complexity has obvious pay-offs. They are easy to apply, and they enable the researcher to make quantitative comparisons between task conditions and between different groups of learners with economy. But by surrendering a more pragmatic sensitivity to individual learners in the interests of economy and generalizability, we run the risk of failing systematically to discriminate between two kinds of learner. Firstly, those learners who are using language which is sufficiently complex to meet their communicative needs, but who are not pushing their output (or not to the degree they could) for the specific purposes of language learning. Secondly, those learners who are using language which is not merely complex in reference to linguistic categories, but which is also complex relative to their own abilities – output, in other words, which is more rigorously and deliberately being pushed towards a learning goal. How the current bias towards quantitative data might be redressed in order to be more sensitive to learner differences is a difficult question. At the very least, though, there must be a case for using think-alouds and other protocol analyses to systematically probe individual learner's experience of the planning process, both strategic and on-line (see Ortega in Chapter 3 and Sanguran in Chapter 4).

Planning as skilled discourse

Planning as meta-cognitive skill

In this section we will examine various aspects of planning, which tend to be regarded from a cognitive viewpoint as unproblematic, but which from a discourse perspective are seen as relatively complex, demanding and/or ambiguous.

Effective planning is a skilful and demanding activity requiring a careful, conscious and selective engagement with language (Skehan 1998b: 88–91; O'Malley & Chamot 1990). As such, planning constitutes a complex of skills requiring an ability to think abstractly about language and about related meaning, and strategic planning (in particular) requires in addition a capacity to think about one's linguistic and discourse intentions in relation to the discourse context, including how one might anticipate or respond to the needs of interlocutors. Yet despite this complexity, planning *per se* in the SLA literature is not considered to be a particular challenge. Indeed, it is standard practice to launch learners quite peremptorily into whatever planning tasks are the focus of the study, without taking any prior soundings about their experience of planning or their capacity to engage with it effectively.

Ortega (whose 1999 study of planning is a rare case of combining quantitative data analysis with a discussion of learners' strategic interpretations of the planning process) gives the following learner's account of planning:

I corrected those verbs, even though I'm no gonna have the notes to do the exercise, but it still helped me to know, that's how I remember things, by writing them, so. Cause writing even though I never got to look at it again, it helped me to remember how to say things ... (1999: 129)

It is standard procedure in most of the planning studies to provide learners with paper on which they can make notes as they plan, but then to take these notes away before they commence their 'planned' performance. Yet the effectiveness of this apparently simple procedure is underpinned by an ability to abstract meta-cognitively about one's relevant linguistic knowledge, and to be able to retain at least some of this plan and subsequently act on it. Ortega's learner, one could argue, demonstrates a considerable level of meta-cognitive awareness in terms of which learning strategies do or do not work well for her – an awareness which cannot be assumed to be shared by the general population of second language learners.

The best laid plans: Strategic planning as discourse skill

The ability to use one's meta-cognitive skills to formulate a plan which is commensurate with one's goal is itself an impressive feat. But its effectiveness is entirely dependent on one's ability subsequently to contextualize the strategic plan on-line i.e. to operationalize it as discourse. Even if we leave aside the question of how learners engage with the subsequent task as discourse participants, there remains the sizeable question of how (or whether) they can formulate plans, which are, in principle, practical in terms of being realized in subsequent discourse.

Scholars sometimes argue that the most effective planning may require an ability to separate out planning meaning from planning form. Specifically, this might involve an ability to prioritise the former during strategic planning, whilst reserving detailed linguistic planning for the on-line task (see Ellis 2003: 133). This in itself demonstrates a not inconsiderable level of strategic ability. But beyond that, the ability to hold on to one's plan and to pursue it even when under on-line pressure requires a level of discourse self-awareness and assertiveness which may be difficult to sustain once the task itself (with all its pragmatic uncertainties) is underway, as Skehan and Foster in Chapter 7 have demonstrated. Ortega provides a good example of a learner who, in effect, admits that on-line pressure towards communicative appropriacy made it difficult to retain the original plan:

I noticed, though, when I was writing it, you think more, it's easier to think of it. I don't know why it's easier to think of the grammar, grammar-wise rather than just when you speak. So, when I was speaking, I wasn't really thinking of the grammar. I was just thinking 'oh, God, how do I explain this?', 'what did I say on the paper? I forgot!' (1999: 131)

Discourse relevance and planning for interaction

The ability to integrate plan and performance is of particular concern when the post-task performance requires learners to engage in group interaction. In Foster and Skehan (1999), for example, learners are asked to plan a 'balloon debate': a discussion supposedly set in a hot air balloon which is losing altitude, where participants have to justify why they should be allowed to remain in the balloon rather than be jettisoned in order to save the lives of fellow passengers. This is a task requiring very considerable skill, in particular given the unpredictability of the subsequent debate: how does one plan with any conviction a contribution to a debate whose shape and evolution are largely unknown?

It would be interesting to see how learners might be explicitly assisted in planning for this kind of eventuality by being supported through a period of instruction in collaborative learning and group planning. Amongst other things, they might well benefit from working on ways of framing their discussion so as to allow for their collective plans to be implemented. For example, by allowing each speaker an initial period to lay out their case uninterrupted, and/or by having a discussion stage during which interlocutors can ask questions of clarification but cannot challenge another participants' position in terms of argument, thus helping them to capitalize on their strategic plans whilst also supporting a measured departure from them.

But the kind of support which learners typically get (if any) in the planning studies is generally of a much more rudimentary kind. In one of the group conditions in the Foster and Skehan study, for example, students were guided by the researcher to focus on content rather than form by being given a character to defend and being told to "think together of good reasons why their character should not be thrown out of the balloon". (1999:247). Subsequently the students were re-assembled into new groups, with one student present from each of the four previous groupings, before being allowed to have their debate. Thus the potential effectiveness of whatever plans individual learners had made may well have been sacrificed, made vulnerable to the need to work in new groupings where the compatibility of individual's plans would be down to chance. Short of entirely disregarding even the most basic notions of discourse relevance – for example by sticking to the plan irrespective of whatever else was being said in the group – such learners would have surely found it difficult (if not impossible) to use their plans in an effective way. Indeed, any attempt to remain faithful to a prior plan might simply have rendered the learners less pragmatically alert than they might have been had they never attempted to plan in the first instance.

Underrating the learners' potential agency

For these reasons we should not be surprised to discover that in Foster and Skehan's study, group planning consistently produced much less effective results than planning which was explicitly directed by the teacher, and that in terms of complexity there was little to choose between the group planners and the control group who were allowed no opportunity to plan at all (1999:235). Yet the researchers imply that this may be a reflection of the superiority of teacher controlled planning over planning which is learner initiated. Teacher-led planning, they say, "... is likely to introduce a greater level of efficiency

to all learners since it is the product of preparation on a teacher's part, and a greater degree of organization." (1999:223). But one wonders whether planning which is learner-initiated would be quite so ineffective where participants are allowed to take more direct advantage of their planning, and where they are not so hamstrung (as they may have been in Foster and Skehan's study) by being disoriented through re-grouping, and by an absence of discourse-sensitive planning guidance.

Planning, socio-history and educational culture

Planning is just one of a number of strategies for learning located within a progressivist educational philosophy, a philosophy premised on the belief that learners should shoulder greater responsibility for managing their own learning (O'Malley & Chamot 1990). Some learners, however, will have had more prior experience of this kind of educational culture than others, and the learners used in the planning studies often appear to be of this sort: university teaching assistants (in the case of Williams 1992); university students (in Ortega's 1999 study) and so on. On the assumption that learners' personal and educational histories exert a strong influence on their present learning potential (Donato 2000), one wonders whether learners in many of the planning studies benefit disproportionately from an experience of relatively western educational contexts and higher levels of educational achievement compared (say) to refugee second language learners coming from more teacher-centred educational systems (see the argument in Block 2003:32–58).

Nor is it simply a matter of learners either being good planners or not. Some learners' backgrounds might predispose them towards planning for one kind of linguistic outcome over another (e.g. accuracy over fluency, or accuracy over complexity), as suggested by Ortega in Chapter 3. Crookes (1989), noting that ten minutes of strategic planning time had relatively little impact on his Japanese learners, speculates on whether discourse spontaneity may not be something commonly found among Japanese, and "thus a predisposition towards the use of planning (both co-planning and pre-planning) may have limited the differences which might otherwise have been seen." (1989:379). Here, though, we should be cautious about over-essentializing the relationship between learners' cultural histories and their pedagogic behaviours.

Communicative constraints or learning discourse: The arguments in principle

This section examines the main rationale which planning studies provide for why learners might choose to push their output. This rationale is a communicative one: learners push their language in order to meet the informational needs of their interlocutor, in the guise either of a fellow learner who will be physically present during the post-planning task, or a fictional interlocutor whose communicative needs are suggested in the task rubric. It is this kind of transactional discourse motivation which I refer to using the more general term 'communicative'; in other words, a discourse based on the need to focus on and exchange information made relevant by a particular task.

The equation between communicative discourse and output pushing is questionable, however, because a sensitivity to one's interlocutor can easily deter output pushing and lead one instead towards conservatism and linguistic reductionism. What might be needed, rather, is a very different kind of discourse, referred to here as a 'learning discourse', in which learners are encouraged to take risks with their output as the primary goal, even where such risks threaten conventional communicative parameters on relevance and acceptability. In this section the focus is on the arguments which underpin communicative and learning discourses in principle. In the following section data from recent planning studies are re-examined in order to provide tentative evidence of both these discourses in action.

Discourse motivation in SLA: Communicative need or explicit learning purpose?

Much of the research into planning has focused on the interaction between strategic planning and task design, and speculation about the impact of task design features on learner performance, with considerable effort invested into looking for predictive relationships between the cognitive and linguistic complexity of task design features and learners' language in subsequent planned performance. The results of such studies certainly show a greater-than-chance overall correlation between task specifications and learner output, and the results for complexity are said to be more consistent than they are either for fluency or for accuracy (Ellis 2003: 131/132).

But there are reasons for applying caution when interpreting such results. One reason for caution has to do with the kind of motivation which drives learners to plan. In order to push their output learners will likely need to se-

lect, consciously and strategically, 'cutting edge' linguistic forms, which they then integrate into the resulting discourse. In mainstream SLA, the most effective means for facilitating such a process are thought to be communicative. In the planning studies it is common for learners to be told that they will need to bear in mind the communicative needs of their interlocutor (e.g. Crookes 1989; Williams 1992; Foster & Skehan 1996; Mehnert 1998; Ortega 1999). If the task involves describing pictures, for instance, the learners may be asked to pay special attention to the clarity and orderliness of what they say because their interlocutor's task is to correctly arrange and sequence the pictures on this basis (e.g. Ortega 1999; Gilanlioglu 2001).

Ortega argues that the planning process encourages learners to operationalize a principle for attending to language form which is central to much SLA scholarship: focus on form, conceived of as a kind of engagement with language which requires a meaningful, non-mechanistic use of form through prior engagement with meaning (see Long 1991; Long & Robinson 1998). Making the link between focus on form and communicative need explicit, Ortega argues that a 'stringent definition' of focus on form must presuppose that "the targeted form(s) be an integral part of the communicative needs engendered by the instructional language use event" (1999: 110). She goes on to argue that this type of focus on form "may result in increased opportunities for making form-function connections, for noticing the gap . . . for noticing holes in one's competence . . . and, in sum, for restructuring and development" (*ibid.*).

We need, though, to be a little careful when considering such a hypothesis. It may well be that the kind of language required on grounds of communicative appropriacy is at odds with the kind of language required for output pushing and related processes of progressive automatization, and for one of two reasons. Firstly, output pushing often requires taking discourse risks (such as the risk of being unclear and thus of losing face). Such factors may well constrain the learner towards linguistic conservatism and away from output pushing in the interests of communicative clarity. Secondly, the very fact that learners are alerted to their interlocutor's agenda may make them so communicatively careful that they disregard all but the most overtly relevant information. Indeed, in such circumstances some learners may regard the opportunity to plan strategically as an opportunity not to stretch their linguistic resources but rather to ensure that their output is strictly warranted on grounds of communicative economy. As will be suggested shortly, underpinning these two kinds of discourse conservatism are two closely related principles of discourse motivation: the principle of clarity (requiring the speaker to be informative and clear) and

the principle of economy (requiring the speaker to be brief and economical: see Grice 1975; Leech 1983; Poulisse 1997).

Different discourses: Communicative discourse and learning discourse

A learning discourse differs fundamentally from communicative discourses, and particularly from transactional/communicative discourses, which are predominantly the focus of SLA research. In communicative discourse, linguistic form is used as the main vehicle towards attaining a communicative end (Widdowson 1983), just as in 'focus on form' language is elaborated as "an integral part of . . . communicative needs" (Ortega 1999:110). In short, form is the means whilst meaning (and meaning clarification) is the end. But in a learning discourse deliberately shaped to facilitate output pushing, the direction of dependency is reversed: form is the end, and meaning is the means.

We can usefully think of these as discourse interpretations, and even as interpretations which different learners might bring to bear on the same task (cf. Breen 1985; Coughlan & Duff 1994). We might consider, as an example, a task where learners are asked to plan a narrative based on a picture sequence. Interpreted communicatively, the essential task might be to provide a description of sufficient clarity to enable the listener to sequence the pictures correctly (even if this is at the cost of taking significant risks with one's language). Interpreted as a learning discourse, the essential task might be to use the pictures as a point of departure for selecting language that is personally challenging to the learner (even if this is at the cost of the interlocutor's need for communicative clarity).

Matters of degree: Interpretation, agency and submissiveness

It is not being suggested that learners face a dichotomous choice here: either enact a communicative purpose and lose out as learners, or disregard communicative norms, push your output for its own sake and consequently prosper as learners. Undoubtedly there is a level of communicative performance at which learners can use language which is complex – even pushed to a degree – whilst also bearing in mind their interlocutor's needs. But even here, communicative appropriacy will likely have a constraining effect on a learners' output relative to the facilitating effects of a learning discourse. As was suggested in section two, however, the planning studies in SLA are unable to distinguish between these two – between communicative learners whose language is sufficiently complex to register as such, and learners who go much further by pushing their language specifically for the sake of developing it (as with a learning discourse).

Finally, it is important to note that many learners who interpret planning along broadly communicative lines may do so not as assertive and self-directing agents, but in a rather more submissive and reactive mode. Many learners respond to the experience of operating in a second language by experiencing an enhanced sense of face threat (Aston 1986), coupled with a loss of their own 'voice' (Pavlenko & Lantolf 2000). This tendency towards self-effacement is entirely congruent with a cautiously communicative approach to planning. It is arguable, indeed, that communicative conditions provide the perfect camouflage for face sensitive learners who exploit principles of clarity/economy to the hilt out of a concern with self-protection rather more than with self-expression.

Communicative constraints or learning discourse: Some tentative evidence

Clarity, comprehensibility and risk in communicative discourse

The need for clarity in discourse has long been recognized (it is famously encapsulated by Grice (1975) as his maxim of manner: 'be clear'), and many of the planning studies feature task instructions which point very clearly in this direction (see Sanguran, Chapter 4, for a discussion of the importance of task rubrics in planning). The instructional rubric provided to learners in Ortega's (1999) planning task is a case in point. Here the learners are told that their partner "needs to get the complete story from you" and that they should "try . . . to be clear and specific, so that your partner can find out about her/his pictures" (1999: 147). Ortega goes on to quote some of the learners in her study who express an awareness of the link between their partner's needs and their own focus on linguistic clarity; one talks of the need not to 'throw the person off', whilst another wants to avoid 'messing up the person I'm telling the story' (1999: 128).

What is equally significant about such rubrics is the absence of any counter-balancing orientation towards interlanguage stretching. Hardly ever in the planning studies are learners specifically encouraged to think about their language development needs, and only in relatively few cases are they explicitly directed towards the use of language form (but see Hulstijn & Hulstijn 1984; Foster & Skehan 1999).

One might object to this argument, however, noting that SLA researchers are in fact very aware of the importance of form and of the need for learners to use planning to engage more deeply with forms which are more challenging

and ‘cutting edge’. The very notion of complexity is defined by some scholars in just these terms. Foster and Skehan (1996:304) note that complexity “is likely to be associated with greater risk-taking to the extent that actual performances may be exploited to use forms closer to the cutting edge of interlanguage development”.

But whilst a correlation between complexity and risk taking is often asserted, it is not so often explained or argued for. Complexity is of its nature a linguistic concept (as was noted earlier on), whereas risk taking is essentially a matter of discourse. Risk, in other words, implies that one engages in some form of behaviour notwithstanding the potential cost, and it is through discourse that cost and loss are made real: loss of face, loss of confidence, loss of clarity. One could, then, take issue with some SLA scholars for raising the question of risk taking without following through its discourse implications.

We need to think much more carefully about how to encourage learners to take risks and not to interpret the discourse context primarily in terms of communicative economy. With this in mind, Ellis (1987) provides a small segment of data which may be instructive. In this data, we find a learner who is charged with writing a story on the basis of picture prompts. The fact that this is a written (and not an oral) account, as Ellis argues, gives the learner greater attentional space within which she can organize her linguistic output:

The thief and his young collaborator *had taken* a car and *had disappeared* on the traffic. They stopped on a forest but they *had haven* a big surprise . . .

(Ellis 1987:10. Italics added)

Speculatively, there are reasons for thinking that this learner might be interpreting the task as a learning discourse. Part of Ellis’s procedure involved asking the learners to begin their narrative with the words ‘One day . . .’, the rationale being to “encourage the use of past tense forms” (1987:6). That this particular learner might have taken this as a prompt to take risks with her syntax is suggested by the very noticeable use (indeed, over use) of the italicised past and past perfect forms. Again speculatively, this is the kind of pushed output we might expect from a learner who is experimenting with novel forms, testing out a new hypothesis by playing with particular form/function combinations (see Swain 1995), or simply gaining experiential practice in accessing cutting edge grammar.

What might add substance to this version of events is the fact that in Ellis’s task (as in many other planning studies) there is no interlocutor, neither physically present nor suggested in any sense in the task rubric. Yet far from needing an interlocutor in order to orient their use of language, these learners appear

to thrive on its absence (see also Bygate 1996). Note how the language in Ellis's data is potentially very unclear at certain points. For example, did the 'big surprise', which the narrative's protagonists experience occur before or after their arrival in the forest? There is a marked ambiguity here traceable directly to the learner's risk-taking with the past perfect and past participle ('they *had haven* a big surprise . . .'). One can only wonder whether this learner would have felt so prepared to threaten comprehensibility in this way had she been addressing and shaping her language to and for an interlocutor. The use of interlocutor-free monologues of this kind may be exactly the sort of un-communicative, pedagogic contrivance which is needed in order to encourage learners systematically to take risks with their output (see Batstone 2002b).

Redundancy and economy in a communicative discourse

Scholars in SLA have for long recognized the potential for learners to exploit shared knowledge so as to cut down on what they need to say (Bolinger 1975; Swain 1985). This kind of linguistic conservatism is particularly prevalent in communicative tasks where learners are able to capitalize on context to avoid spelling out information which they have good reason to suppose is easily inferred from context. For example, had the learner in Ellis's study interpreted the task in a more communicative/economical way, she would have felt no need to use past tense forms of any kind: the researcher has already given her the contextual prompt 'one day', rendering any subsequent grammaticization of pastness redundant.

Unless overtly alerted to the contrary, some learners might see the opportunity to plan as an opportunity not to avoid economy but rather to capitalize upon it. In other words, they might use their planning time carefully to sort out the strictly relevant from the strictly irrelevant. Not all tasks in the planning literature lend themselves to such an approach, but many (such as picture ordering) certainly do. Ortega quotes one learner in her picture-sequencing task whose account of the planning process is strongly oriented towards economy:

I tried to think about *what information she would need*, so I was thinking more of her [during the preparation time]. It gave me time to think about her perspective, rather than just try to get the story out. (Ortega 1999: 129. My italics)

Gilanglioglu (2001) provides us with a great deal of data where we can see his learners working their way through picture sequencing tasks which, like those used by Ortega, include in their rubric a clear direction towards communicative clarity. "Because the pictures are quite similar to each other", the rubric states,

“your partner needs an accurate description as possible in order to decide on the correct order of the pictures” (2001:250). Here is the transcript of one such task, in which Cem (C: the picture describer) is taking Umut (U: the interlocutor) through a set of photographs of an office. Umut’s task is to distinguish one picture from the next and assign each picture its correct number:

- C: Can you see the masks?
 U: Yes
 C: Okay. In picture one, there’s no mask in the sight area.
 U: Okay
 C: Okay?
 U: Yes
 C: In picture two and three, there’s a white mask, but in picture two the face of the mask doesn’t look to me
 U: Okay
 C: Okay?
 U: Okay
 C: in picture three, the white mask looking to me
 U: Yes, okay
 C: Okay?
 U: Uh-huh

(Gilanlioglu 2001:324)

What is striking about Cem’s performance is the degree to which he has interpreted the task in terms of communicative economy. The idea that economy as a discourse principle might enable learners to cut down (sometimes quite dramatically) on their output is hardly new (see, for example, Seedhouse 1999). But what seems to be happening in Cem’s case is not so much a wholesale abandonment of grammatical elaboration, but rather a more measured approach where the opportunity to plan is being used to pinpoint just those specific details necessary to distinguish one picture from another. And if we check his description of each picture against the original, we can see the precision with which he has gone about his task (nor is his precision unwarranted, since the pictures in this task differ only in matters of small detail).

Looking through Gilanlioglu’s transcriptions, it would probably be fair to say that Cem’s approach to the task is a particularly scrupulous one. Nonetheless, his economy (and his resulting linguistic brevity) serve to indicate how planning which is not carefully directed towards the need to take risks can be interpreted in quite other directions by the learner, and at some pedagogic cost.

Redundancy, economy and relevance in a learning discourse

Irving (1999) describes a judgement task which is repeated by the same two learners. The second time around, unprompted by the researcher, both learners effectively re-invent the task by converting it into a monologue, and both deploy language of greater complexity. Monologues (as was noted earlier) can free learners from being overly constrained by thoughts of clarity or economy. But in Irving's case (unlike in Ellis 1987 or Bygate 1996) both learners are co-present throughout the task, so that they continue to talk in the presence of an interlocutor (and indeed they do occasionally interrupt each other). Nonetheless, each learner's monologue begins with a quite detailed account of a narrative they are both intimately acquainted with. We would be hard pressed to interpret this as communicative discourse, because virtually all of the information being expressed here is already established as shared knowledge and is thus redundant to the task.

Planning meaning as a means in a learning discourse

In Irving's study we have a good example of a learning discourse in operation. Rather than using form primarily as a means to achieve a meaning-oriented goal (in this case, to rank a set of characters in order of guilt for causing a series of accidents), meaning is used as a device for launching much more richly into linguistic form. Indeed, it would be very difficult to interpret Irving's data communicatively and to regard her learners as using language as a means towards achieving meaning clarification as the primary goal. The narrative story here is already firmly established as a 'given'. Consequently a more plausible explanation for why these learners behaved the way they did is that they deliberately chose to elaborate their language at the very point where (in a communicative discourse) little if anything further would have been said. We cannot possibly understand how learners in this study use freed up attentional resources to push their output without seeing how this whole cognitive process is mediated and motivated through a very particular (and a distinctly non-communicative) interpretation of discourse and discourse relevance.

Another possible example of meaning being used in order better to target pushed linguistic forms is suggested in the study by Coughlan and Duff (1994), who discuss how different learners interpreted the same tasks in divergent ways through weaving in their own particular detail and plot. Such creativity may or may not be evidence of a learning discourse, but it is intriguing to note how it might be. Learners, in other words, deliberately improvise additional mean-

ing not simply to create interest or to engage an interlocutor, but in order to construct a richer propositional network around which to develop their linguistic resources. Once again, then, meaning is a deliberate means towards a linguistic end. Planning, of course, could be used for precisely this purpose: to help learners elaborate meanings to act as hooks upon which to hang their subsequent linguistic explorations. In the planned narrative task in Gilanlioglu's study (2001), many learners added considerably to the richness of their language by improvising content. Here, for example, are some extracts from different learners of a description of the first picture (of a girl travelling in a car en route to a skiing holiday):

Student A

"It was a bright and shiny day. The girl was going to mountain for holiday just to relax and forget about job – her job or city life." (2001:580)

Student B

"I have a friend who has a strange story. She's called Mary and one day she got bored of her life and wanted to have a different holiday..." (2001:588)

Student C

"Mary Jean was a hardworking secretary in Chicago and just for a change she wanted to go skiing to a mountain" (2001:595)

Student D

Dave is a successful manager – in a company and he er got bored one day from business and er he decide he realized that he has he needs a holiday..." (2001:598)

It is worth noting that the task here is a simple one of picture ordering, and none of these details are of any significance in helping the interlocutor to decide on the correct sequence. Thus there is no specific communicative/transactional reason for this kind of creativity. But that, of course, may well be the point, for these learners may be fabricating meaning as a deliberate contrivance towards a language pushing objective. Gilanlioglu notes that the overall results in terms of complexity on the planned narrative task were not of great statistical significance relative to the unplanned narrative. Even so, his data are at least suggestive, and we can certainly speculate about how much more complex and lexically diverse these discourses might have been had the learners been overtly guided to use narrative improvising towards an explicit learning purpose.

Implications and applications

Summary

Current research into planning underplays the complexity of learning and learning discourse. In terms of theoretical framework, the insights of cognitive approaches carry only limited explanatory power on their own, and need to be socio-cognitively situated in a theory which (in one way or another) recognizes the interdependence of cognition, social contexts and discourse interpretation. In terms of their view of language, planning studies over-rely on linguistic measures (as with complexity) and do not explore their discourse implications. In terms of task design, SLA planning research tends to overestimate the predictability of task effects on learner language, and significantly to underplay the mediating role of learner interpretation. In elevating the task, planning research diminishes the significance of the learner. Yet the key to understanding a learning discourse lies as much with learners and their discourse identities as it does with an analysis of tasks and their design features. Indeed, all the key features of a learning discourse outlined above presuppose, above all else, a different inner orientation to language (a different sense of what is relevant in terms of contribution, of what is acceptable in terms of risk). As a corollary, the strongest potential benefit to language pedagogy may lie not so much with task design (though that is still important), but with the developing of a 'learning culture' in classrooms where learners share a common discourse understanding which legitimises risk taking and validates linguistic elaboration for its own sake (see Batstone 2002a).

Implications

Finally, whilst there are grounds for wanting a broader theoretical basis to understand planning and cognition in SLA, there are also grounds for being cautious in framing specific recommendations for change. Much of the foregoing argument in favour of an alternative discourse understanding of planning is based on speculation, and on a favourable interpretation of data from other scholars, which may or may not be justified. What is needed is not merely more data, but a triangulation of different approaches to data collection. Specifically, generalizable but insensitive data of the kind commonly in use in SLA needs to be set against measures which are more sensitive (as was argued earlier), and in two regards. First they need to be more psycholinguistically sensitive, in the sense that they probe the interlanguage boundaries of individual learners

and seek to establish to what degree real output pushing is going on (see Ellis's comments, in this volume, about using planning to target more specific forms through pre and post testing). Secondly, they need to be more sociolinguistically sensitive, and for this we need much more qualitative data, systematically elicited and used to seek out evidence for learner interpretations which we can set alongside more linguistic and quantitative measures.

Such suggestions are easy to make but their implementation is likely to prove complex and challenging. One suspects, though, that the payoff will be worth the investment, and that it will lead us towards a conception of planning which will reveal more about how second language learning processes are sociocognitively shaped.

References

- Allan, D. 1992. *Oxford Placement Test*. Oxford: Oxford University Press.
- Anderson, J. 1983. *Architecture of Cognition*. Cambridge MA: Harvard University Press.
- Anderson, J. 2000. *Learning and Memory: An integrated approach* (Revised Edition). New York NY: John Wiley and Sons.
- Anderson, N. J. and L. Vandergrift. 1996. Increasing metacognitive awareness in the L2 classroom by using think-aloud protocols and other verbal report formats. In R. L. Oxford (ed.), *Language Learning Strategies around the World: Cross-cultural perspectives. Technical Report # 13*. Honolulu HI: University of Hawai'i, Second Language Teaching & Curriculum Center.
- Aston, G. 1986. Trouble-shooting in interaction with learners: The more the merrier. *Applied Linguistics* 7: 128–143.
- Bachman, L. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. 2002. Some reflections on task-based language performance assessment. *Language Testing* 19(4): 453–476.
- Baddeley, A. 1986. *Working Memory*. Oxford: Oxford University Press.
- Baddeley, A. and G. Hitch. 1974. Working memory. In G. Bower (ed.), *The Psychology of Learning and Motivation* (Vol. 8). New York NY: Academic Press.
- Baddeley, A. and R. Logie. 1999. The multiple-component model. In A. Miyake and P. Shah (eds), *Models of Working Memory*. New York NY: Cambridge University Press.
- Batstone, R. 2002a. Making sense of new language: A discourse perspective. *Language Awareness* 11(1).
- Batstone, R. 2002b. Contexts of engagement: A discourse perspective on 'intake' and 'pushed output'. *System* 30: 1–14.
- Bayley, R. 1996. Competing constraints on variation in the speech of adult Chinese learners of English. In R. Bayley and D. Preston (eds), *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins.
- Bereiter, C. and M. Scardamalia. 1987. *The Psychology of Written Composition*. Hillsdale NJ: Lawrence Erlbaum.
- Bialystok, E. 1990. *Communication Strategies: A psychological analysis of second language use*. Oxford: Basil Blackwell.
- Block, D. 2003. *The Social Turn in Second Language Acquisition*. Edinburgh: Edinburgh University Press.
- Bolinger, D. 1975. *Aspects of Language* (2nd Edition). New York NY: Harcourt, Brace, Johanovitch.

- Bourdin, B. and M. Fayol. 1994. Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology* 29: 591–620.
- Breen, M. P. 1985. The social context for language learning – a neglected situation? *Studies in Second Language Acquisition* 7: 135–158.
- Breen, M. P. 2001. Postscript: New directions for research on learner contributions. In M. P. Breen (ed.), *Learner Contributions to Language Learning: New directions in research*, 172–182. New York NY: Pearson Education.
- Brindley, G. and H. Slatyer. 2002. Exploring task difficulty in ESL listening assessment. *Language Testing* 19: 369–394.
- Brock C. 1986. The effects of referential questions on ESL classroom discourse. *TESOL Quarterly* 20(1): 47–59.
- Brown G., A. Anderson, R. Shilcock and G. Yule. 1984. *Teaching Talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.
- Brown, G. 1995. *Speakers, Listeners, and Communication*. New York NY: Cambridge University Press.
- Brown, G. and G. Yule. 1983. *Teaching the Spoken Language*. New York NY: Cambridge University Press.
- Brown J. D., T. Hudson, J. Norris and W. Bonk 2002. *An Investigation of Second Language Task-based Performance Assessments*. Honolulu HI: University of Hawai'i Press.
- Brumfit, C. 1984. *Communicative Methodology in Language Teaching*. Cambridge: Cambridge University Press.
- Buckwalter, P. 2001. Repair sequences in Spanish L2 dyadic discourse: A descriptive study. *The Modern Language Journal* 85: 380–397.
- Butler, F. A., D. Eignor, T. McNamara, S. Jones and B. K. Suomi. 2000. TOEFL 2000 Speaking Framework: A working paper. *TOEFL Monograph Series MS-20*. Princeton NJ: Educational Testing Service.
- Bygate, M. 1996. Effects of task repetition: Appraising the developing language of learners. In J. Willis and D. Willis (eds), *Challenge and Change in Language Teaching*. Oxford: Heinemann.
- Bygate, M. 1999. Task as the context for the framing, re-framing and unframing of language. *System* 27: 33–48.
- Bygate, M. 2001. Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan and M. Swain (eds), *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing*. Harlow: Longman.
- Bygate, M. 2004. Task repetition: Some issues for applied linguistics and language pedagogy. Plenary talk to Asociación Española de Lingüística Aplicada. Valencia, April 2004.
- Bygate, M., P. Skehan, and M. Swain. (eds). 2001. *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing*. Harlow: Longman.
- Chamot, A. U. 2001. The role of learning strategies in second language acquisition. In M. P. Breen (ed.), *Learner Contributions to Language Learning: New directions in research*. New York: Pearson Education.
- Clark H. H. and J. Fox Tree. 2002. Using 'uh' and 'um' in spontaneous speaking. *Cognition* 84: 73–111.

- Clark, H. and G. Clark. 1977. *Psychology and Language: An introduction to psycholinguistics*. New York NY: Harcourt, Brace, Jovanovich.
- Clark, R. 1974. Performing without competence. *Journal of Child Language* 1: 1–10.
- Cohen, A. D., S. J. Weaver, and T.-Y. Li. 1998. The impact of strategies-based instruction on speaking a foreign language. In A. D. Cohen (ed.), *Strategies in Learning and Using a Second Language*. New York NY: Longman.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale NJ: Lawrence Erlbaum.
- Cooper, T. 1976. Measuring written syntactic patterns of second language learners of German. *Journal of Educational Research* 69: 173–183.
- Coughlan, P. and P. A. Duff. 1994. Same task, different activities: Analysis of a SLA task from an activity theory perspective. In J. Lantolf and G. Appel (eds), *Vygotskian Approaches to Second Language Research*. Norwood NJ: Ablex.
- Crookes, G. 1989. Planning and interlanguage variation. *Studies in Second Language Acquisition* 11: 367–383.
- Davies, A. 1990. *Principles of Language Testing*. Oxford: Blackwell.
- De Bot, K. 1992. A bilingual production model: Levelt's 'Speaking Model' adapted. *Applied Linguistics* 13: 1–24.
- De Larios, J., J. Marin and L. Murphy. 2001. A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning* 51: 497–538.
- DeKeyser, R. 1993. The role of practice in second language learning: A perspective from cognitive psychology. Paper presented at the AILA conference, Amsterdam, August 8–13.
- DeKeyser, R., R. Salaberry, P. Robinson and M. Harrington. 2002. What gets processed in Processing Instruction? A commentary on Bill VanPatten's 'Processing Instruction: An update.' *Language Learning* 52: 805–823.
- Dickerson, L. 1975. The learner's interlanguage as a system of variable rules. *TESOL Quarterly* 9: 401–407.
- Donato, R. 2000. Sociocultural contributions to understanding the foreign and second language classroom. In J. Lantolf (ed.), *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.
- Dörnyei, Z. 2002. The motivational basis of language learning tasks. In P. Robinson (ed.), *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins.
- Z., and J. Kormos. 2000. The role of individual and social variables in oral task performance. *Language Teaching Research* 4: 275–300.
- Doughty, C. 2001. Cognitive underpinnings of focus on form. In P. Robinson (ed.), *Cognition and Second Language Instruction*. New York NY: Cambridge University Press.
- Doughty, C. and J. Williams (eds). 1998. *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.
- Ejzenberg, R. 1992. *Understanding Nonnative Oral Fluency*. Ann Arbor MI: U.M.I.
- Elder C., N. Iwashita and T. McNamara. 2002, Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? *Language Testing* 19(4): 343–368.
- Ellis, N. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24: 143–188.

- Ellis R. 1987. Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition* 9: 1–20.
- Ellis, R. 1990. *Instructed Second Language Acquisition*. Oxford: Blackwell.
- Ellis, R. 2003. *Task-based Language Teaching and Learning*. Oxford: Oxford University Press.
- Ellis, R. and G. Barkhuizen. 2004. *Analyzing Learner Language*. Oxford: Oxford University Press.
- Ellis, R. and F. Yuan. 2004. The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies in Second Language Acquisition* 26: 59–84.
- Esser, U. 1995. Oral language testing: The concept of fluency revised. Unpublished MA thesis, Lancaster University.
- Faerch, C., and G. Kasper. 1983. Plans and strategies in foreign language communication. In C. Faerch and G. Kasper (eds), *Strategies in Interlanguage Communication*. Harlow: Longman.
- Fillmore, C. 1979. On fluency. Reprinted in Riggenbach, H. (ed.) 2000. *Perspectives on Fluency*. Ann Arbor MI: University of Michigan Press.
- Firth, A. and J. Wagner. 1997. On discourse, communication, and (some) fundamental concepts in SLA research. *Modern Language Journal* 81: 286–300.
- Flower, L and J. Hayes. 1980. The dynamics of composing: Making plans and juggling constraints. In L. Gregg and E. Steinberg (eds), *Cognitive Processes in Writing*. Hillsdale NJ: Lawrence Erlbaum.
- Foster, P. 1996. Doing the task better: How planning time influences students' performance. In J. Willis and D. Willis (eds), *Challenge and Change in Language Teaching*. London: Heinemann
- Foster, P. 2001. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In Bygate et al. (eds).
- Foster P. and P. Skehan. 1996. The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition* 18(3): 299–324.
- Foster, P. and P. Skehan. 1997. Modifying the task: The effects of surprise, time and planning type on task based foreign language instruction. *Thames Valley Working Papers in English Language Teaching* 4.
- Foster, P. and P. Skehan. 1999. The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research* 3: 215–247.
- Foster, P., A. Tonkyn. and G. Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21: 354–375.
- Freed B. 2000. Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (ed.), *Perspectives on Fluency*. Ann Arbor MI: The University of Michigan Press.
- Gass, S. M. and A. Mackey. 2000. *Stimulated Recall Methodology in Second Language Research*. Mahwah NJ: Lawrence Erlbaum.
- Gass, S., A. Mackey, M. Fernandez and M. Alvarez-Torres. 1999. The effects of task repetition on linguistic output. *Language Learning* 49: 549–580.
- George, H. V. 1972. *Common Errors in Language Learning*. Rowley MA: Newbury.
- Gilanlioglu, I. 2001. The effects of task features on lexis and grammar in L2 oral performance. Institute of Education, University of London, PhD thesis (draft).
- Givon T. 1985. Function, structure, and language acquisition. In D. Slobin (ed.), *The Cross-Linguistic Study of Language Acquisition*. Hillsdale NJ: Lawrence Erlbaum.

- Glaser, B. and A. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for qualitative research*. Chicago IL: Aldine.
- Grabe, W. 2001. Notes on a theory of second language writing. In T. Silva and P. Matsuda (eds), *On Second Language Writing*. Mahwah NJ: Lawrence Erlbaum.
- Grabe, W. and W. Kaplan. 1996. *Theory and Practice of Writing: An applied linguistic perspective*. Harlow UK: Longman.
- Grice, P. 1975. Logic and conversation. In P. Cole and J. Morgan (eds), *Syntax and Semantics, Vol. 3: Speech acts*. New York NY: Academic Press.
- Hatch, E. 1983. *Psycholinguistics: A second language perspective*. Rowley MA: Newbury House.
- Hatch, E. and A. Lazaraton. 1991. *The Research Manual*. Boston MA: Heinle and Heinle.
- Hayes, J. and L. Flower. 1980. Identifying the organization of the writing process. In L. Gregg and E. Steinberg. (eds), *Cognitive Processes in Writing*. Hillsdale NJ: Lawrence Erlbaum.
- Hayes, J. and J. Gradwohl Nash. 1996. On the nature of planning in writing. In C. Levy and S. Ransdell (eds), *The Science of Writing*. Mahwah NJ: Lawrence Erlbaum
- Heaton, J. 1975. *Beginning Composition through Pictures*. London: Longman.
- Hill, L. A. 1960. *Picture Composition Book*. London: Longman.
- Hirano, K. 1991. The effect of mode of discourse on objective measures of EFL proficiency in Japanese university students. *JACET Bulletin* 22: 15–34.
- Hoey M. 1983. *On the Surface of Discourse*. London: Allen and Unwin.
- Huitt, W. 2003. The information processing approach. <http://chiron.valdosta.edu/whuitt/col/cogsys/infoproc.html>
- Hulstijn, J. 2002. Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second language Research* 18: 193–223.
- Hulstijn, J. and W. Hulstijn. 1984. Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning* 34: 23–43.
- Irving, R. 1999. An investigation into social and psychological effects in oral interactive tasks. Unpublished MA thesis. University of London, Institute of Education.
- Iwashita, N. 2001. The effect of learner proficiency on interactional moves and modified output in nonnative-nonnative interaction in Japanese as a foreign language. *System* 29: 267–287.
- Iwashita N., C. Elder and T. McNamara. 2001. Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning* 51(3): 401–436.
- Kaiser H. 1974. An index of factorial simplicity. *Psychometrika* 39: 31–36.
- Kamimoto, T. and C. Kawauchi. 2000. The effects of repetition on EFL oral fluency: A quantitative approach. *Kumamoto Gakuen University Journal* 7: 1–24.
- Kawauchi, C. 1998. Cued/uncued oral narrative tasks in foreign language fluency. *JACET Bulletin* 29: 83–96.
- Kawauchi, C., and T. Kamimoto. 2000. Distinctive features of oral production by fluent and nonfluent EFL learners. *Language Laboratory* 37: 17–32.
- Kayfetz, J. L. and R.L. Stice. 1987. *Academically Speaking*. Boston MA: Heinle and Heinle.
- Kellog, R. 1996. A model of working memory in writing. In C. Levy and S. Ransdell (eds), *The Science of Writing*. Mahwah NJ: Lawrence Erlbaum.

- Kirsch I. and P. Mosenthal. 1990. Exploring document literacy: variables underlying performance of young adults. *Reading Research Quarterly* 25(1): 5–30.
- Kobayashi, M. 1995. The effects of text structure and test format on the measurement of reading comprehension. Unpublished Ph.D. thesis, Thames Valley University.
- Kobayashi, M. 2002. Method effects on reading comprehension test performance: Text organization and response format. *Language Testing* 19(2): 193–220.
- Koponen, M. and H. Riggenbach. 2000. Overview: Varying perspectives on fluency. In H. Riggenbach (ed.), *Perspectives on Fluency*. Ann Arbor MI: The University of Michigan Press.
- Kormos J. and Denes M. (2004), “Exploring measures and perceptions of fluency in the speech of second language learners”, *System*, 32, 145–164.
- Krashen, S. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.
- Labov, W. 1970. The study of language in its social context. *Studium Generale* 23: 30–87.
- Lantolf, J. 1996. Second language theory building: Letting all the flowers bloom! *Language Learning* 46: 713–749.
- Lantolf, J. P. (ed.). 2000. *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.
- Laufer, B. and P. Nation. 1995. Lexical richness in L2 written production. *Applied Linguistics* 16: 307–322.
- Leech, G. 1983. *The Principles of Pragmatics*. London: Longman.
- Levelt, W. 1989. *Speaking: From intention to articulation*. Cambridge MA: The MIT Press.
- Linacre, M. 1990. *FACETS Computer Program for Many Faceted Rasch Measurement*. Chicago IL: Mesa Press.
- Lindemann, S. 2002. Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society* 31: 419–441.
- Lockhart, C. and P. Ng. 1995. Analyzing talk in ESL peer response groups: Stances, functions, and content. *Language Learning* 45: 605–655.
- Long M. 1988. Instructed interlanguage development. In L. Beebe (ed.), *Issues in Second Language Acquisition: Multiple perspectives*. Rowley MA: Newbury House.
- Long, M. H. 1991. Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg and C. Kramersch (eds). *Foreign Language Research in Cross-Cultural Perspective*. Amsterdam: John Benjamins.
- Long, M. H. and G. Crookes. 1993. Unit of analysis in syllabus design: The case for task. In G. Crookes and S. M. Gass (eds), *Tasks in a Pedagogical Context: Integrating theory and practice*, (Vol. 1). Clevedon: Multilingual Matters.
- Long, M. and P. Robinson. 1998. Focus on form: Theory, research and practice. In C. Doughty and J. Williams (eds), *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.
- Loschky, L. and R. Bley-Vroman. 1993. Grammar and task-based methodology”. In Crookes, G. and S. Gass (eds), *Tasks in a Pedagogical Context: Integrating theory and practice*. Clevedon: Multilingual Matters.
- Lynch, T. and J. Maclean. 2000. Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research* 4: 221–250.

- Lynch, T. and J. Maclean. 2001. Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan and M. Swain (eds).
- Mackey, A., S. Gass and K. McDonough. 2000. Do learners recognize implicit negative feedback as feedback? *Studies in Second Language Acquisition* 22: 471–497.
- MacWhinney, B. 1995. *The CHILDES Project: Tools for analyzing talk*. [2nd edition.] Hillsdale NJ: Lawrence Erlbaum.
- MacWhinney, B. 2001. The competition model: The input, the context, and the brain. In P. Robinson (ed.), *Cognition and Second Language Instruction*. New York NY: Cambridge University Press.
- Malvern, D. and B. Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19: 85–104.
- McLaughlin, B. 1987. *Theories of Second Language Acquisition*. London: Arnold.
- McLaughlin, B. 1990. Restructuring. *Applied Linguistics* 11(2): 113–128.
- McNamara, T. 1996. *Measuring Second Language Performance*. London: Longman.
- Meara, P. 2004. Modelling vocabulary loss. *Applied Linguistics* 25(2): 137–155.
- Mehnert U. 1998. The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition* 20: 52–83.
- Meisel, J., H. Clahsen and M. Pienemann. 1981. On determining developmental stages in natural language acquisition. *Studies in Second Language Acquisition* 3: 109–135.
- Miyake A. and P. Shah. 1999. *Models of Working Memory*. Cambridge: Cambridge University Press.
- Morris, F. A. and E. Tarone. 2003. Impact of classroom dynamics on the effectiveness of recasts in second language acquisition. *Language Learning* 53: 325–368.
- Morrow, K. 1979. Communicative language testing: Revolution or evolution? In C. Brumfit and K. Johnson (eds), *The Communicative Approach to Language Teaching*. Oxford: Oxford University Press.
- Nemeth, N. and J. Kormos. 2001. Pragmatic aspects of task-performance: The case of argumentation. *Language Teaching Research* 5: 213–240.
- Norris, J., J. D. Brown and T. Hudson. 2000. Assessing performance on complex L2 tasks: Investigating raters, examinees and tasks. Paper presented at the Language Testing Research Colloquium, Vancouver, March.
- Norris, J. and L. Ortega. 2001. Does type of instruction make a difference? Substantive findings from a meta-analytic review. In R. Ellis (ed.), *Form-focused Instruction and Second Language Learning*, 157–213. Malden MA: Blackwell.
- Norris, J. M., J. D. Brown, T. Hudson and W. Bonk. 2002. Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing* 19(4): 395–418.
- Nunan, D. 1991. Communicative tasks in the language classroom. *TESOL Quarterly* 25: 279–295.
- Nunnally, J. C. 1967. *Psychometric Theory*. New York NY: McGraw Hill.
- O'Malley, J. M. and A. U. Chamot. 1990. *Learning Strategies in Second Language Acquisition*. New York: Cambridge University Press.
- Ochs, E. 1979. Planned and unplanned discourse. In T. Givon (ed.), *Syntax and Semantics* Vol 12: *Discourse and semantics*. New York NY: Academic Press.

- Ogawa, T. 1990. *Eigoshikaku no sankann-o* [Three best English proficiency tests]. Tokyo: Sanshusha.
- Ortega, L. 1995. The effect of planning in L2 Spanish narratives. *Research Note 15*. Honolulu HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Ortega L. 1999. Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition* 21: 109–148.
- Ortega, L. In press. Meaningful L2 practice in foreign language classrooms: An SLA perspective." In R. DeKeyser (ed.), *Practice in Second Language Learning: Perspectives from Linguistics and Psychology*. New York NY: Cambridge University Press.
- Oxford, R. L. 1990. *Language Learning Strategies: What every teacher should know*. New York NY: Newbury House/Harper Collins.
- Pavlenko, A. and J. Lantolf. 2000. Second language learning as participation and the (re) construction of selves. In J. Lantolf (ed.).
- Pellettieri, J. 2000. Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer and R. Kern (eds), *Network-based Language Teaching: Concepts and practice*. New York NY: Cambridge University Press.
- Pica, T. 1983. Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition* 6: 69–79.
- Pica, T. 1994. Research on negotiation: What does it reveal about second-language learning condition, processes, and outcomes? *Language Learning* 44: 493–527.
- Pica, T. 1997. Second language teaching and research relationships: A North American view. *Language Teaching Research* 1: 48–72.
- Platt, E. and F. B. Brooks. 2002. Task engagement: A turning point in foreign language development. *Language Learning* 52: 365–400.
- Pollitt A. 1990. Giving students a sporting chance: Assessment by counting and judging. In C. J. Alderson and B. North (eds), *Language Testing in the 1990's* [Modern English Publications]. London: The British Council.
- Poullisse, N. 1997. Compensatory strategies and the principles of clarity and economy. In G. Kasper and E. Kellerman (eds), *Communication strategies: Psycholinguistic and sociolinguistic perspectives*, 49–64. London: Longman.
- Prabhu, V. S. 1987. *Second Language Pedagogy*. Oxford: Oxford University Press.
- Preston, D. 1996. Variationist perspectives on second language acquisition. In R. Bayley and D. Preston (eds), *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins.
- Quirk R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.
- Rampton, B. 1997. Second language research in late modernity: A response to Firth and Wagner. *Modern Language Journal* 81: 329–333.
- Ranta, L. 2002. The role of learners' language analytic ability in the communication classroom. In P. Robinson (ed.), *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins.
- Raupach M. 1980. Temporal variables in first and second language production. In H. W. Dechert and M. Raupach (eds), *Temporal Variables in Speech: Studies in honour of Frieda Goldman-Eisler*. The Hague: Mouton.

- Regan, V. 1996. Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley and D. Preston (eds), *Second Language Acquisition and Linguistic Variation*, 177–201. Amsterdam: John Benjamins.
- Richards, J., J. Platt and H. Platt. 1996. *Dictionary of Language Teaching & Applied Linguistics*. London: Longman.
- Riggenbach, H. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes* 14: 423–441.
- Robinson, P. 1995. Task complexity and second language narrative discourse. *Language Learning* 45: 99–140.
- Robinson P. 2001a. *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.
- Robinson P. 2001b. Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.
- Robinson, P. 2001c. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 21(1): 27–57.
- Robinson, P. 2003. Attention and memory during SLA. In C. Doughty and M. H. Long (eds.), *Handbook of Second Language Acquisition*. Malden MA: Blackwell.
- Rutherford, K. 2001. An investigation into the effects of planning on oral production in a second language. Unpublished Masters dissertation, Institute of Language Teaching and Learning, University of Auckland.
- Samuda, V. 2001. Guiding relationships between form and meaning during task performance. In M. Bygate, P. Skehan and M. Swain (eds).
- Sangarun, J. 2001. The effects of pre-task planning on foreign language performance. Unpublished doctoral thesis, University of Toronto, Canada.
- Sato, C. J. 1988. Origins of complex syntax in interlanguage development. *Studies in Second Language Acquisition* 10: 371–395.
- Schmidt, R. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11: 129–158.
- Schmidt, R. 1994. Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review* 11: 11–26.
- Schmidt, R. 2001. Attention. In Robinson, P. (ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.
- Schmidt, R. and S. Frota. 1986. Developing basic conversational ability in a second language. In R. Day (ed.), *Talking to Learn*. Rowley MA: Newbury House.
- Seedhouse, P. 1999. Task-based interaction. *English Language Teaching Journal* 53: 149–156.
- Skehan P. 1996. A framework for the implementation of task-based instruction. *Applied Linguistics* 17(1): 38–62.
- Skehan, P. 1998a. Task-based instruction. *Annual Review of Applied Linguistics* 18: 268–86.
- Skehan P. 1998b. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skehan P. 2001. Tasks and language performance. In M. Bygate, P. Skehan and M. Swain (eds).
- Skehan, P. 2002. Theorising and updating aptitude. In P. Robinson (ed.), *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins.

- Skehan P. 2003. Task based instruction. *Language Teaching* 36(1): 1–14.
- Skehan, P. and P. Foster. 1995. Some type and task processing conditions as influences on foreign language performance. *Thames Valley University Working Paper in English Language Teaching* 3: 139–188.
- Skehan, P. and P. Foster. 1997. Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research* 1: 185–211.
- Skehan P. and P. Foster. 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning* 49(1): 93–120.
- Skehan, P. and P. Foster. 2001. Cognition and tasks. In P. Robinson (ed.), *Cognition and Second Language Instruction*. New York NY: Cambridge University Press.
- Spada, N. 1997. Form-focussed instruction and language acquisition: A review of classroom and laboratory research. [State of the art article] *Language Learning* 30(2): 73–97.
- StatView [Computer software]. 1996. (Version 4.5). Berkeley CA: Abacus Concepts.
- Stein, W. 1991. *Communicative Skills that Work: A functional approach for life and work*. Book. Chicago IL: Contemporary Books.
- Stenhouse, L. 1975. *An Introduction to Curriculum Research and Development*. London: Heinemann.
- Storch, N. 2002. Patterns of interaction in ESL pair work. *Language Learning* 52: 119–158.
- Swain, M. 1985a. Large-scale communicative language testing: A case study. In Y. P. Lee, A. Fok, R. Lord and G. Low (eds), *New Directions in Language Testing*, 35–46. Oxford: Pergamon.
- Swain, M. 1985b. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass and C. Madden (eds), *Input in Second Language Acquisition*. Cambridge MA: Newbury House.
- Swain, M. 1993. The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review* 50: 158–164.
- Swain, M. 1995. Three functions of output in second language learning. In G. Cook and B. Seidhofer (eds), *For H. G. Widdowson: Principles and practice in the study of language*. Oxford: Oxford University Press.
- Swain, M. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (ed.), *Sociocultural theory and second language learning*. New York NY: Oxford University Press.
- Swain, M. 2001. Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing* 18(3): 275–302.
- Swain, M. and S. Lapkin. 2000. Task-based second language learning: The uses of the first language. *Language Teaching Research* 4: 251–274.
- Tabachnik B. G. and L. S. Fidell. 1996. *Using Multivariate Statistics*. (3rd Edition). New York NY: Harper Collins.
- Tannen, D. 1984. Spoken and written narrative in English and Greek. In D. Tannen (ed.), *Coherence in Spoken and Written Discourse*. Norwood NJ: Ablex.
- Tarone, E. 1982. Systematicity and attention in interlanguage. *Language Learning* 32: 69–82.
- Tarone, E. 1983. On the variability of interlanguage systems. *Applied Linguistics* 4: 143–63.
- Tarone, E. 1985. Variability in interlanguage systems: A study of style-shifting in morphology and syntax. *Language Learning* 35: 373–403.

- Tarone, E. and B. Parrish. 1988. Task-related variation in interlanguage: The case of articles. *Language Learning* 38: 21–44.
- Thomas, M. 1994. Assessment of L2 proficiency in second language acquisition research. *Language Learning* 44: 307–336.
- Ting, S. C. 1996. Planning time, modality and second language task performance: Accuracy and fluency in the acquisition of Chinese as a second language. *The University of Queensland Working Papers in Language and Linguistics* 1: 31–64.
- Tomlin, R. and V. Villa. 1994. Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition* 16: 183–203.
- Tono, Y. 2000. A computer learner corpus-based analysis of the acquisition order of the English grammatical morphemes. In T. McEnery and L. Burnard (eds), *Rethinking Language Pedagogy: Papers from the third International Conference on Teaching and Language Corpora*. Frankfurt: Peter Lang.
- Towell R., R. Hawkins and N. Bazergut. 1996. The development of fluency in advanced learners of French. *Applied Linguistics* 17(1): 84–115.
- VanPatten, B. 1990. Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition* 12: 287–301.
- VanPatten, B. 1996. *Input Processing and Grammar Instruction*. Norwood NJ: Ablex.
- VanPatten, B. 2002. Processing instruction: An update. *Language Learning* 52: 755–803.
- Wendel, J. 1997. Planning and second language narrative production. Unpublished doctoral dissertation, Temple University, Japan.
- Wenden, A. L. 2001. Metacognitive knowledge in SLA: The neglected variable. In M. P. Breen (ed.), *Learner Contributions to Language Learning: New directions in research*. New York NY: Pearson Education.
- Whalen, J. and N. Menard. 1995. L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning* 45: 381–418.
- Widdowson, H.G. 1983. *Learning Purpose and Language Use*. Oxford: Oxford University Press.
- Wigglesworth G. 1997. An investigation of planning time and proficiency level on oral test discourse. *Language Testing* 14(1): 85–106.
- Wigglesworth, G. 2000. Issues in the development of oral tasks for competency-based assessments of second language performance. In G. Brindley (ed.), *Studies in Immigrant English Language Assessment* (Vol. 1) [Research Series 11]. National Centre for English Language Teaching and Research Macquarie University: Sydney.
- Wigglesworth G. 2001. Influences on performance in task based oral assessments. In Bygate M., P. Skehan and M. Swain (eds).
- Williams, J. 1992. Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly* 26: 693–711.
- Williams, J. 1999. Learner-generated attention to form. *Language Learning* 49: 583–625.
- Winter E. 1976. Fundamentals of information structure: A pilot manual for further development according to student need. The Hatfield Polytechnic Linguistics Group, School of Humanities.
- Wolfe-Quintero, K., S. Inagaki and H. Kim. 1998. *Second Language Development in Writing: Measures of fluency, accuracy and complexity*. Hawaii HI: University of Hawaii Press.

- Yuan, F. and R. Ellis. 2003. The effects of pre task and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics* 24(1): 1–27.
- Yule, G. 1996. *Referential Communication Tasks*. Mahwah NJ: Lawrence Erlbaum.
- Yule, G. and M. Powers. 1994. Investigating the communicative outcomes of task-based interaction. *System* 22: 81–91.
- Yule, G., Powers, M. and D. Macdonald. 1992. The variable effects of some task-based learning procedures on L2 communicative effectiveness. *Language Learning* 42: 249–277.
- Zimmerman, R. 2000. L2 writing: Subprocesses, a model of formulating and empirical findings. *Learning and Instruction* 10: 73–99.

Index

A

- accuracy 14–19, 22–26, 28–32,
79–82, 84, 91–93, 112–114, 120,
123–127, 129–131, 140, 143–146,
153–162, 164–166, 168–171,
185–191, 194–198, 220–222,
227–234, 243–248, 254–264,
266–272,
- accuracy oriented learners 93
- activity (vs. task) 1, 3, 63, 67, 75,
145, 146, 149, 150, 155, 160, 162,
163, 197, 221, 236, 277, 281
- agency 283, 287
- articulation 13, 14, 41, 43, 44, 124,
173, 174, 189, 254
- AS unit 204, 256, 257
- attention 5–14, 17, 22–24, 34,
40–42, 44, 67, 75–78, 93,
100–102, 104–107, 111, 128–131,
162, 163, 170, 193–197, 212–214,
240–241, 243, 246, 278, 286
- automaticity 14, 254, 277, 280

B

- breakdown fluency viii, 32, 199,
203, 205, 211, 217, 254, 266–268

C

- c-units (C-units) 32, 123, 125, 181,
215, 230, 256, 279
- careful planning (*see also* on-line
planning and within-task
planning) 4, 11, 169, 173, 175,
176, 179, 180, 188–190, 217
- cognitive models 6, 15
- communication oriented learners 92

- communicative constraints 285, 288
- communicative stress 83, 117, 143,
145, 220
- communicative language teaching
37
- complexity 23–27, 30–32, 75–77,
79–82, 89–91, 112–114, 154–157,
159–161, 164–166, 184–191,
203–209, 220–223, 227–234,
243–245, 256–269, 279–281,
283–285, 289, 292–294
- conceptualization 11, 12, 13, 17, 27,
119, 129–131, 172, 224

D

- declarative knowledge 41, 42, 219
- detailed strategic planning 208, 211
- discourse motivation 285, 286
- disfluency 81, 205

E

- economy (of effort) 19, 275, 280,
286–292
- error-free clauses 32, 123, 168, 182,
198, 211, 228, 256
- exemplar-based knowledge 15, 16

F

- FACETS 220, 227, 229
- fluency 14–16, 18–28, 30–32, 40, 42,
82, 112–114, 154–157, 159–161,
170, 181, 183, 184, 186–191, 194,
199–200, 201, 203–208, 220–222,
227–230, 232–234, 245–248, 254,
255, 257–259, 266–271

focus-on-form(s) 6, 9, 10, 33, 106, 107, 193, 194
 form-focused strategic planning 112–114, 116, 125–127, 141
 formulation 12–14, 17, 22, 24, 33, 39, 41, 43–46, 55, 60, 64, 68, 129–131, 171–173, 174, 187, 188, 190, 196, 210, 213, 220, 246, 269, 270
 framing 35, 46–66, 70, 283, 294

I
 individual differences 63, 78, 84, 91, 93, 94, 105
 information processing 1, 6, 219, 222
 integrative planning 37, 38, 42, 43, 68
 interlanguage vii, 8, 9, 11, 15–17, 90, 191, 212, 275, 277, 288, 289, 294
 interview 18, 81, 82, 87, 104, 105, 107, 108, 122, 180, 187, 188, 202

L
 language proficiency (*see also* proficiency) 13, 31, 33, 34, 76, 79, 143, 164, 171, 173, 175, 191, 235, 248, 251, 258, 263
 language-testing 25, 66, 217, 220, 222, 223, 225, 236, 239, 241–243
 learning discourse 279, 285, 287–289, 292, 294
 length of planning time 21, 22, 234
 length of run 32, 199, 203–208, 210, 254, 255, 258, 259, 261–264, 267, 268
 Levelt 11–15, 22, 41, 43, 44, 118, 119, 121, 172–174, 196, 213, 246, 270
 lexical variety 32, 40, 113, 184–186, 188, 189, 191
 lexico-grammatical knowledge 41, 44, 45, 48–54, 59, 62, 63, 65, 68

limited capacity 6, 14, 33, 106, 167, 174, 240
 listeners' needs 98, 105, 117, 121

M
 macro-planning 12, 13, 39, 40, 172
 meaning-focused strategic planning 112, 113, 116, 119, 125–127, 141
 measuring learner production 30
 metacognition 78, 82, 84, 86
 micro-planning 12, 13, 40, 172, 173, 188
 monitoring 10, 13, 14, 17, 19, 22, 25, 33, 41, 43, 44, 64, 84, 85, 95–98, 101–103, 105, 129, 130, 173–175, 187–191, 196, 205, 213
 monologic tasks 240, 244

N
 narrative task 20–23, 25, 105, 108, 143, 146–148, 152, 155, 165, 196, 226, 243, 247, 248, 255, 276, 293
 negotiation of form 105
 noticing 6, 7, 9, 37, 101, 146, 178, 286

O
 Output Hypothesis (*see also* pushed output) 7, 17, 146

P
 planned language use (*see also* planned discourse) 4, 7
 planning instructions 29, 30, 117–119, 122, 150, 216
 planning time 20–23, 25, 26, 40, 41, 44, 47, 65, 87, 91, 94, 101, 113, 114, 144, 145, 170, 171, 180, 191, 194, 195, 197, 200, 202, 203, 210, 214, 217–225, 228–236, 241, 243, 244, 247, 250–252, 256, 257, 263, 266, 268, 284, 290
 pre-task planning 1, 3–5, 8, 11, 14, 17, 18, 23, 25, 27, 29, 30, 33, 34,

75, 77, 78, 81–85, 87, 88, 90, 91,
94–97, 101, 104–107, 121, 122,
241, 243, 248, 252, 256, 257, 263,
265, 266, 268, 269, 273, 8
pressured planning 176, 179, 180,
187, 188
procedural knowledge 175, 187, 219
proficiency (*see also* language
proficiency) 13, 31, 34, 33, 132,
143, 144, 160, 164, 171, 173, 175,
176, 177, 188
pushed output (*see also* Output
Hypothesis) 277, 279, 280, 289

Q

questionnaire 76, 87, 149, 150, 163,
180, 187, 188, 224–226, 228, 229,
232, 233, 235, 251, 253, 265

R

rate of speech 151, 266, 268
rating(s) 18, 26, 30, 31, 164, 171,
221, 222, 224, 227, 229, 232, 233,
236, 237, 239, 241, 242, 269, 271,
272
rehearsal 3, 4, 8, 14, 18, 19, 23, 28,
29, 33, 35, 85, 87, 89, 94, 95, 97,
104, 105, 143, 145, 146, 149, 150,
163, 164, 219, 235, 241
relevance 5, 7, 33, 43, 78, 94, 235,
243, 246, 282, 283, 285, 292
repair fluency 31, 165, 199, 203,
206–208, 217, 255, 258, 259, 266,
268
repetition 3, 18, 35, 38, 42–46, 48,
49, 51–53, 60, 64–69, 92, 145,
149, 151, 199, 235, 255, 258, 262,
264, 267, 268
resource-depleting factors 16
resource-directing factors 16
retrieval 85, 88–91, 94, 95, 97, 100,
104, 105, 145, 171
rule-based knowledge 15, 16, 25,
220

S

socio-cognitive orientation 275, 276
speech production 6, 11–13, 39, 41,
43, 46, 118, 121, 172–174, 196,
213, 219, 222–224, 235, 246, 254
speech rate 20, 82, 124, 127, 151,
222, 254, 255, 258, 259, 262–264,
267, 268
strategic planning 3–5, 19–25, 29,
38–42, 65, 75, 76, 111–119,
124–127, 129–132, 133–142,
146–148, 164, 190, 194–197,
200–203, 207–214, 217, 221, 223,
229–232, 247, 248, 250–251,
263–265, 268–270, 280–282
strategies (planning) viii, 14, 15, 29,
33, 83–87, 94–98, 101, 104, 105,
108, 117, 145, 146, 219, 233, 234,
242, 281, 284
structured tasks 249, 258, 261, 263,
265–267, 269, 270
stylistic variation 6, 10, 15
surprise information 193, 200, 201,
209–211

T

T-units 32, 158, 181, 215, 256, 279
task characteristics 194, 220, 235,
240, 241, 243, 244, 255, 270, 271
task complexity 16, 90, 178, 220, 223
task conditions 171, 179, 180, 214,
220, 255, 269, 270, 280
task design 1, 23, 27, 28, 30, 33, 34,
46, 78, 80, 81, 111, 115, 116, 148,
176, 191, 192, 196, 223–225, 234,
236, 248, 250, 252, 253, 275, 285,
294
task difficulty 87, 217, 220, 221,
223–225, 228, 229, 231, 232, 241,
251, 265, 269, 270
task implementation 1, 34, 47, 81,
131, 295
task instructions 139, 178, 233, 288

testing context 18, 25–27, 217, 218,
247

trade-off 23, 24, 145, 189, 191, 194,
223

U

undetailed strategic planning 112,
113, 165, 209, 211

unplanned discourse 4

unstructured task 26, 217, 221, 249,
261, 265, 266, 269

V

variability 10, 169, 170, 191

W

within-task planning (*see also* careful
planning and on-line planning)
vii, 3–5, 8, 14, 165, 167–170, 172,
173, 175, 176, 179, 181, 186–188,
190

working memory 6, 8–10, 13, 14,
17, 39, 40, 44, 45, 167, 171, 173,
174, 193, 196, 213

writing 6, 7, 11, 13–15, 31, 32, 39,
76, 78, 81, 85, 87–89, 91, 94, 95,
98, 102, 103, 117, 122, 143, 145,
146, 148–150, 152, 153, 163–165,
167, 168, 170–181, 183–192, 281,
282, 289

In the series *Language Learning & Language Teaching* the following titles have been published thus far or are scheduled for publication:

- 1 **CHUN, Dorothy M.:** *Discourse Intonation in L2. From theory and research to practice.* 2002. xviii, 285 pp. (incl. CD-rom).
- 2 **ROBINSON, Peter (ed.):** *Individual Differences and Instructed Language Learning.* 2002. xii, 387 pp.
- 3 **PORTE, Graeme Keith:** *Appraising Research in Second Language Learning. A practical approach to critical analysis of quantitative research.* 2002. xx, 268 pp.
- 4 **TRAPPES-LOMAX, Hugh and Gibson FERGUSON (eds.):** *Language in Language Teacher Education.* 2002. vi, 258 pp.
- 5 **GASS, Susan M., Kathleen BARDOVI-HARLIG, Sally Sieloff MAGNAN and Joel WALZ (eds.):** *Pedagogical Norms for Second and Foreign Language Learning and Teaching. Studies in honour of Albert Valdman.* 2002. vi, 305 pp.
- 6 **GRANGER, Sylviane, Joseph HUNG and Stephanie PETCH-TYSON (eds.):** *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching.* 2002. x, 246 pp.
- 7 **CHAPELLE, Carol A.:** *English Language Learning and Technology. Lectures on applied linguistics in the age of information and communication technology.* 2003. xvi, 213 pp.
- 8 **JORDAN, Geoff:** *Theory Construction in Second Language Acquisition.* 2004. xviii, 295 pp.
- 9 **SCHMITT, Norbert (ed.):** *Formulaic Sequences. Acquisition, processing and use.* 2004. x, 304 pp.
- 10 **BOGAARDS, Paul and Batia LAUFER (eds.):** *Vocabulary in a Second Language. Selection, acquisition, and testing.* 2004. xiv, 234 pp.
- 11 **ELLIS, Rod (ed.):** *Planning and Task Performance in a Second Language.* 2005. viii, 312 pp.