SSLA, **26**, 59–84. Printed in the United States of America.
DOI: 10.1017/S0272263104261034

THE EFFECTS OF PLANNING ON FLUENCY, COMPLEXITY, AND ACCURACY IN SECOND LANGUAGE NARRATIVE WRITING

Rod Ellis
University of Auckland

Fangyuan Yuan
University of Pennsylvania

Building on previous studies of the effects of planning on second language (L2) learners' oral narratives and drawing on Kellog's (1996) model of writing, this article reports a study of the effects of three types of planning conditions (pretask planning, unpressured on-line planning, and no planning) on 42 Chinese learners' written narratives elicited by means of a picture composition. The results show that, whereas pretask planning resulted in greater fluency (syllables per minute, $p < .01$) and greater syntactic variety (number of different verb forms, $p < .01$), the opportunity to engage in unpressured on-line planning assisted greater accuracy (error-free clauses, $p < .05$). It is proposed that the two types of planning impact on different aspects of L2 writing processes, with pretask planning promoting formulation and unpressured on-line planning providing better opportunities for monitoring. Writers in the no-planning condition were faced with the need to formulate, execute, and monitor under pressure, with negative consequences for the fluency, complexity, and accuracy of the written product in comparison to the planning groups.

This study draws on two different but potentially commensurable research areas: task-based research in SLA, and first (L1) and second (L2) language writing. Task-based research has been primarily concerned with the effects of task

The authors would like to thank the anonymous *SSLA* reviewers of an earlier version of this article for their perceptive and constructive criticism.

Address correspondence to: Rod Ellis, Department of Applied Language Studies and Linguistics, University of Auckland, Private Bag 92019, Auckland, New Zealand; e-mail: r.ellis@auckland.ac.nz.

© 2004 Cambridge University Press 0272-2631/04 $12.00

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

design and implementational variables on the fluency, complexity, and accuracy of language in oral production. Psycholinguistically oriented writing research has typically used data collected from think-aloud protocols to identify the strategies used by writers and to model the mental systems responsible for the production of written text. Clearly, these are very different traditions. However, as Kellog (1996) noted, it is reasonable to assume that processes involved in oral and written production have much in common. Thus, we maintain, much is to be gained by drawing on insights from both research areas.

TASK-BASED RESEARCH: THE EFFECTS OF PLANNING

A number of studies have investigated the effects of planning on L2 learners' performance of oral narratives (e.g., Ellis, 1987; Foster & Skehan, 1996; Ortega, 1999; Robinson, 1995; Skehan & Foster, 1997, 1999; Wendel, 1997; Yuan & Ellis, 2003). These studies showed that giving learners the opportunity to plan a narrative before they speak it (i.e., pretask planning) resulted in significant gains in both fluency (whether measured in terms of temporal variables such as number of syllables per minute or hesitation variables such as frequency of reformulations) and complexity (measured most commonly in terms of the degree of subordination). However, these studies produced mixed results when the focus was accuracy, as measured, for example, by the percentage of error-free clauses. Although Ellis found that pretask planning led to increased accuracy in the use of regular past-tense verbs in oral narratives in English, Wendel found no effect on accuracy in Japanese learners' narrative productions. Other studies have also produced mixed results where accuracy is concerned. For example, Ortega found that pretask planning led to greater accuracy in the use of noun modifiers in L2 Spanish but not in the use of articles. Overall, these studies demonstrated that pretask planning aids fluency and complexity but not necessarily accuracy in L2 learners' oral narratives.

Although a number of studies have investigated the effects of pretask planning, only a few studies have examined on-line planning (i.e., the planning that occurs during a speech event). Drawing on Levelt's (1989) model of speaking, Yuan and Ellis (2003) suggested that, when learners have the time to plan what they are going to say, they use this time to both formulate their message more carefully and monitor the output of the formulation and articulation phases of production. Both operations involve greater attention to form. Two other studies have also examined this type of planning. Hulstijn and Hulstijn (1984) asked learners of L2 Dutch to perform short oral narratives under four conditions that involve combinations of two variables: time (i.e., the subjects 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 either form or 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. Yuan and Ellis found that, like pretask planning, on-line planning

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

(operationalized as unpressured performance) resulted in increased complexity of oral language use. However, whereas pretask planning did not result in greater accuracy in L2 learners' narrative production, on-line planning did, although no opportunity for pretask planning was provided. These results suggest that L2 learners of limited proficiency are only able to improve the accuracy of their oral production when they are focused on form and have sufficient time to formulate what they want to say and monitor their output.

Yuan and Ellis's (2003) results support the contention that L2 performance is characterized by competing demands, such that attention to one aspect of language detracts from learners' ability to attend to another aspect. Foster and Skehan (1996) proposed that the trade-off is between accuracy and complexity. In contrast, Wendel (1997) proposed that pretask planning promotes fluency, whereas on-line planning enhances accuracy. Thus, he saw the trade-off as involving fluency and accuracy. Yuan and Ellis's study lended greater support to Wendel's position. Thus, whereas both types of planning result in enhanced complexity, pretask planning aids fluency but not accuracy, whereas on-line planning has a detrimental effect on fluency (as might be expected) but improves accuracy.

In contrast to the number of studies that have investigated the effects of planning on oral narratives, there have been very few task-based studies of the effects of planning on written narratives. Ellis (1987) examined the accuracy levels of a mixed group of learners' use of three English past-tense forms (regular, irregular, and copula) in a written narrative completed without any time pressure (thereby allowing for on-line planning). He found that overall accuracy levels were higher in this task than in two oral tasks (one with and one without pretask planning) but that only the difference for regular past tense was statistically significant. In other words, on-line planning assisted grammatical accuracy in the written narrative but only in the case of a feature for which there was a simple, "portable" rule. Ellis did not investigate fluency or complexity.

WRITING RESEARCH: THE ROLE OF PLANNING

In writing research, planning has been viewed as one of several processes involved in the production of written text. Its role, therefore, needs to be considered in relation to the other composing processes. This has been addressed through models of the complete writing process. Cumming (in press) pointed out that there is no universally accepted theory of L2 writing. Nor, as Grabe (2001) pointed out, are there any specifically L2 theories of writing. However, the available theories (e.g., Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; Grabe; Grabe & Kaplan, 1996; Kellog, 1996; Zimmerman, 2000) posit a very similar set of processes, and there is general acceptance that these are broadly similar in both L1 and L2 writing.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

We have chosen Kellog's (1996) model as the basis of this study. This model draws heavily on earlier models of writing (e.g., Flower & Hayes, 1980) in positing three basic processes involved in writing. However, it also extends the earlier models by relating these processes to Baddeley's (1986) model of working memory. Kellog's model is limited in that it does not take account of the role of the task environment nor of affective factors in writing (see Hayes's, 1996, extension of the original Flower and Hayes model for an ambitious attempt to incorporate these factors). However, this model fits our purposes well in that our study is not designed to investigate the effects of task environment, and it affords a relatively simple framework that is nevertheless capable of explaining the complex nature of writing processes.

The model (shown diagrammatically in Figure 1) distinguishes three basic systems involved in text production. Each system has two principal components or processes. *Formulation* entails planning, during which the writer establishes goals for the writing, thinks up ideas related to these goals, and organizes these to facilitate action, and translating, when 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 programming, where the output from translation is converted into production schema for the appropriate motor system involved (e.g., handwriting or typing), and executing or the actual production of sentences. *Monitoring* consists of reading, where the writer reads

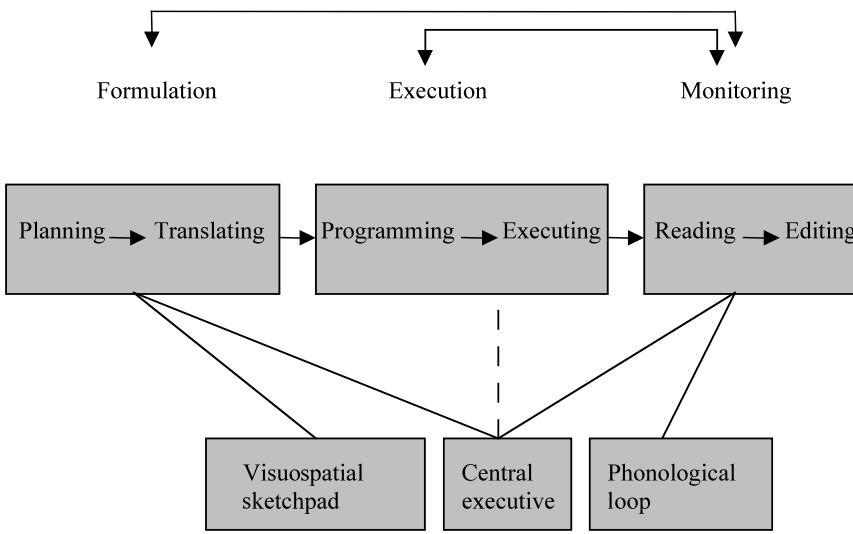


Figure 1. Kellog's model of writing processes. Adapted with permission from: Kellog, R. (1996). A model of working memory in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing* (p. 59). Mahwah, NJ: Erlbaum.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

his or her own text ("a necessary but not sufficient condition for writing well"; Kellog, 1996, p. 61), and 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, macro aspects such as paragraph and text organization, or both aspects. The extent to which a writer is able to engage in monitoring depends in part on whether the writer has the time to adopt a polished draft strategy or is engaged in pressured text production, as in Elbow's (1981) freewriting. Kellog, like the authors of similar models, emphasized that writers simultaneously activate formulation, execution, and monitoring processes although the extent to which this activation is achievable depends on working memory.

Kellogg (1996) detailed how the different components of the model relate to different components of working memory. He argued that the central executive, a multipurpose system responsible for problem solving, mental calculation, and reasoning, is involved in all the subprocesses with the exception of executing, which, he argued, is usually accomplished without the need for controlled processing. It should be noted, however, that this assumes an adult, nativelike automaticity in handwriting or typing, which may not be present in young children (see, e.g., Bourdin & Fayol, 1994) or in some L2 learners, especially those whose L1 employs a different script. It is possible, therefore, that the central executive may be called on by some L2 writers during execution. Kellog suggested that the visuospatial sketchpad, which stores and processes visual and spatial information in working memory, is only involved in planning. Finally, he proposed that the phonological loop, which stores and processes auditory and verbal information, is required for both translating and reading. The key feature of Kellog'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 prioritize when under pressure to produce text rapidly. This is reflected in a trade-off of attention directed at the different processes. Previous research (e.g., Brown, McDonald, Brown, & Carr, 1988) has suggested that formulation demands are critical and take priority over execution and monitoring.

It is clear that Kellog's (1996) model of writing is very similar to Levelt's (1989) model of speech production. Thus, Kellog's "planning" corresponds to Levelt's "conceptualization," "translating" is equivalent to Levelt's "formulation," and "execution" to "articulation." Both models recognize the role of monitoring and also acknowledge that it can influence all the other processes. In one crucial respect, however, writing differs from speech production. Speaking has to be accomplished in real time, whereas writing allows the writer much greater control over the time spent in formulation and monitoring. In the case of L2 learners with limited proficiency, this may be crucially important. Because speaking typically entails pressure to perform in real time, it may lead to greater levels of anxiety in L2 learners than writing, with the obvious exception of fast writing.[1]

Although designed to account for L1 writing, Kellog's (1996) model is also applicable to L2 learners. Nevertheless, it is important to ask in what ways L2

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

writing differs from L1 writing. Differences are likely to arise as a result of the increased pressure on working memory that L2 writers experience due to limited proficiency. For example, learners who have limited L2 linguistic resources or difficulty in accessing these resources are likely to find translating problematic. De Larios, Marin, and Murphy (2001) found that L2 writers, especially those with a lower level of proficiency, need to concentrate on translation at the expense of on-line planning and revising. In other words, they attend more to form. They explained this by suggesting that the central executive inhibits some processes while activating others to ensure completion of the task. Such differences are best seen as quantitative rather than qualitative in nature.

There is, however, one potential qualitative difference between L1 and L2 writing. This concerns the L2 writer's use of the L1. However, Zimmerman (2000) and Raab (1992, as cited by Zimmerman) found little evidence of L2 writers using their L1 during translation, although they did report its use for global planning to assist lexical searches and as an aid during editing. Woodhall (2002) reported that how often L2 writers used their L1 depended on their proficiency, with advanced learners relying on it less than intermediate learners. In contrast, the length of language switches was related to task difficulty, with the more difficult of the two tasks leading to the use of the L1 for longer periods of time. This study failed to find a linear relationship between use of the L1 and text quality for the writers as a whole, but in the case of writers whose L1 was cognate with the L2, use of the L1 assisted performance in the more difficult of the two tasks. The L1 is best seen as a resource that learners draw on variably.

We now consider in greater detail how planning has been defined in writing research and how these definitions relate to the distinction between pretask and on-line planning. Hayes and Gradwohl Nash (1996) saw planning as a type of reflection, to be considered with other reflective processes such as decision making and inferencing (as in Hayes's, 1996, revised model). However, they argued that planning can be distinguished from other types of reflection in that it occurs in an environment different from the environment of the task itself. They were clearly thinking of what we have called pretask planning. They distinguished two types of such planning. Process planning is focused on the writer and how the task is to be performed (i.e., identifying strategies for accomplishing the task). Text planning is focused on the content and form of what is to be written. This can entail abstract text planning that leads to the production of ideas, notes, and outlines involving content and rhetorical organization, language planning, or both. Hayes and Gradwohl Nash also recognized what we have called on-line planning by acknowledging that planning and text production are often interwoven. They defined *construction tasks* as "tasks that produce their own output gradually with considerable interleaving of plans and action, and the output influences the subsequent planning" (p. 41). Whether planning occurs inside or outside the task environment is, then, the distinguishing feature of on-line and pretask planning.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

Whalen and Menard (1995) discussed planning in terms of the discourse levels involved. *Pragmatic planning* involves defining pragmatic objectives (i.e., identifying audience and a personal or professional reason to write and detailing the topic content). *Textual planning* entails determining an appropriate text typology and how to achieve coherence between idea sequences. Finally, *linguistic planning* occurs when the writer attempts to solve a linguistic problem to formulate an idea. In their study, they did not make a clear distinction between pretask and on-line planning, with the implication that the three types of planning and the strategies associated with them apply to both.

There have been a number of studies of the effects of pretask planning. Hayes and Gradwohl Nash (1996) reviewed a number of L1 writing studies. They concluded that "the effect of planning on text quality is almost entirely attributable to time on task" and that "planning is neither more nor less valuable than other writing activities" (p. 53). In other words, pretask planning leads to better quality texts and greater fluency in writing simply because it affords learners more time overall. This conclusion contradicts Grabe's (2001) assertion that "a writer who plans for no more than 10 seconds will write a predictable essay with less information (and most likely a lower quality essay) than a student who plans for 4 minutes" (p. 40). It should be noted, however, that many of the studies Hayes and Gradwohl examined employed holistic ratings of writing as measures of text quality rather than measures of information content or of specific linguistic and discourse variables.

There have been few studies of the effects of pretask planning on L2 writing. To address the question "What does time buy?" Kroll (1990) compared the written compositions produced in class by L2 writers with a 60-minute time limit with those produced at home. She found that the learners produced more accurate language and more highly rated writing in the compositions produced at home, but the difference was not statistically significant. As Kroll recognized, a problem with this study is that there was no way of knowing how long the writers spent on the compositions written at home or whether the two conditions resulted in different planning behaviors. Friedlander (1990) explored the hypothesis that L2 writers plan their writing more effectively and write better texts containing more content when they are able to plan in the language in which they acquired information about a topic. This hypothesis was supported. However, the plans and essays in the matched and unmatched conditions, overall, received identical ratings regardless of the language they planned in.

Whalen and Menard (1995) compared the types of planning used in an argumentative writing task in their L1 and L2. They found that pragmatic and linguistic planning was more likely to occur in their L1 writing but that there was no difference in the quality of their planning in the two languages. Their study used think-aloud protocols. Like all such studies that have investigated the processes involved in on-line assembly using this method (e.g., Zimmerman, 2000), planning seemed to have afforded writers plenty of time to compose; thus, their study can be assumed to have investigated within-task

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

planning. De Larios, Marin, and Murphy (2001) also investigated within-task planning. They found that formulation (i.e., translation in Kellog's, 1996, model) played a dominant role in the production of an argumentative text in relation to other composing processes such as planning. The L2 writers with greater proficiency, however, were better able to attend to planning. These studies do not allow clear conclusions regarding the role that planning plays in L2 writing. There is some evidence that it matters little whether the pretask planning is done in the L1 or L2 unless knowledge of the topic was initially acquired through the L1. There is also some evidence that the attention L2 writers need to pay to translation may detract from their ability to engage in planning conceptual content on-line.

In general, then, in the case of L2 writing, the role that either pretask or on-line (i.e., within task) planning plays in the production of written text remains unclear. This is in part because of the failure to make a clear distinction between the two types of planning and also because research to date has tended to rely on holistic ratings rather than the kind of specific linguistic measures that have figured in SLA task-based research. In this study, we explore the differential effects of the two types of planning on the fluency, complexity, and accuracy of L2 written texts.

RESEARCH QUESTIONS

This study was designed to answer the following research questions:

1. Do L2 learners produce more fluent, complex, and accurate language when they have the opportunity to plan a written narrative in advance than when they have no such opportunity?
2. Do L2 learners produce more fluent, complex, and accurate language when they have time for careful on-line planning than when they have limited time?
3. What type of planning (pretask planning or careful on-line planning) has the greater effect on the fluency, complexity, or accuracy of L2 learners' production of written narratives?

We hypothesize that the two types of planning will result in attention being assigned to different components of the composing process with corresponding effects on textual output. The opportunity for pretask planning will encourage process and text planning, including, in some cases, language planning. Thus, writers who have time for pretask planning can be expected to benefit in fluency (defined by Chenoweth & Hayes, 2001, as the rate of production of text) and also in complexity (i.e., they will be able to encode more complex ideas). However, pretask planning is less likely to contribute to translation, execution, and monitoring and thus may have limited effect on accuracy. In contrast, the provision of time for on-line (within task) planning will assist translation, execution, and, in particular, monitoring. It will likely have an adverse effect on fluency, but it may lead to fewer dysfluencies (e.g., false starts

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

and corrections) as writers will be able to iron out linguistic problems before execution. On-line planning time may also foster accuracy but at the price of complexity. Of course, on-line planning must occur in all written production, and the extent to which writers plan on-line is a matter of degree. The comparison intended here is between the effects of pretask planning and the "careful" on-line planning made possible by ensuring that the participants are under no pressure to write quickly.

METHOD

Design

This study is a single-factor, between-participants design with three levels of planning conditions (no planning, pretask planning, and on-line planning). Forty-two participants were administered a pretest to ensure that the three groups had equivalent English proficiency at the outset of the study. Each group produced a written narrative elicited by means of a set of related pictures in one of the three conditions. Their textual products were then analyzed in terms of fluency, complexity, and accuracy.

Participants

The participants were 42 full-time undergraduate English majors in the International Business Department of a Chinese university (i.e., foreign language learners of English)—a departure from other planning studies to date, which have focused on "second" rather than "foreign" language learners. The participants in this study had learned their English more or less entirely in an instructed setting. They were between 18 and 20 years old. At the time of data collection, most of them had been learning English as a foreign language in Chinese schools for 8 years, first in elementary school and middle school and then in college. None 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 were between 100 and 120 (maximum possible = 150), with grades between A and B+ in the oral component of this examination.[2] The participants can be considered to constitute a fairly homogeneous group in terms of their learning history and English proficiency.

As college students, they had 6 hours of English per week—4 hours for reading and writing and 2 hours for listening and speaking. Every 2 weeks, they had a 1-hour oral English class with a native speaker of English from Canada. All the students in two first-year classes were invited and agreed to participate in the study. They were told that the test and tasks 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

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

value. They were not told the precise purpose of the study and were assured that the information collected would not impact their course grades.

The students were divided randomly into three groups of 14 each. The gender composition of each group was as follows: The no-planning group consisted of 8 males and 6 females, the pretask-planning group consisted of 6 males and 8 females, and the on-line-planning group consisted of 5 males and 9 females. No participants withdrew from the study.

Pretest Material

The pretest material was a version of the Test of English as a Foreign Language (TOEFL): Test 1 from *Reading for TOEFL Workbook* (Educational Testing Service, 1987). 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, $F(1, 2) = 0.39$; $p = .95$, or listening scores, $F(1, 2) = 0.464$; $p = .63$. Thus, it can be concluded that the three groups were equivalent in their English proficiency.[3] Descriptive statistics are shown in Table 1.

Task

The task required participants to write a story based on a set of six pictures from Heaton (1975). A written narrative task was chosen to permit comparison with the results of studies that have investigated the effects of planning

Table 1. Descriptive statistics for TOEFL scores across groups

Group scores	*M*	*SD*
On-line planning		
Total	446.71	35.34
Listening	42.79	3.68
Pretask planning		
Total	447.78	27.84
Listening	42.36	3.65
No planning		
Total	460.86	26.57
Listening	43.64	3.46

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

on similar oral tasks (e.g., Foster & Skehan, 1996; Wendel, 1997). We sought to ensure that the task was reasonably demanding on the participants and thus would stretch their linguistic resources. Previous research on oral tasks (e.g., Skehan & Foster, 1999) has indicated that this can be achieved by selecting a picture story that requires interpretation on the part of the learners.

The story 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 for home without noticing the lost package. A man picked up the package and ran 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. This was the same picture story that Ellis (1987) used. The picture story, although clearly structured with a chronologically ordered series of events, requires interpretation on the part of the learners because it is not initially clear that the man had picked up the package, and thus his motive for following the boy is uncertain until the final picture.

The task instructions were given in Chinese. All the participants were given the same prompt to establish the narrative genre required by the task: "This afternoon, Tom. . . ."

Task Conditions

In this study, planning was operationalized at three levels: no planning (NP), pretask planning (PTP), and on-line planning (OLP). The participants performed the task in their normal classroom setting. Both their regular teacher and the researcher were present.

In the NP condition, participants were required to finish the task within 17 minutes and were asked to write at least 200 words. This was intended to limit the amount of time for extensive on-line 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 that study, no time limit was set, and the participants' different times were noted. The pilot study had established that the fastest writer completed the story in 17 minutes, whereas the slowest took 24 minutes. Thus, this condition required "speeded" writing.

In the PTP condition, as in the NP condition, participants were requested to finish writing the story within 17 minutes and to produce at least 200 words. In this way, the participants were pressured to perform the task with limited opportunities for on-line planning. However, in this condition they were given 10 minutes to plan their performance of the task. The choice of planning time was based on Crookes (1989), Foster and Skehan (1996), and Wendel (1997). Mehnert's (1998) study showed that only when at least a 10-minute planning time was provided were there measurable effects on all three aspects of language use—fluency, accuracy, and complexity—in the case of oral production. No detailed guidance was provided, but the participants were asked to

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

plan their narratives in terms of content, organization, and language. This again followed the studies of Crookes, Skehan and Foster, and Wendel. The participants were given a sheet of paper to write notes but told not to write out the whole story. The notes were taken away before they started the task. This achieved two purposes. First, removing the notes ensured that the language elicited by all the tasks was produced within the specified time limit. Second, the notes could be used as evidence regarding how individual students undertook the planning and could be referred to in the posttask interview.

In the OLP condition, the participants were given a piece of paper and told to write down the story. 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 other two groups. Unlike the other two groups, however, the participants were not required to write a minimum of 200 words, as this may have been interpreted as requiring them to write quickly. Thus, the participants in this condition were allocated no time for pretask planning but ample time for on-line planning. The task conditions are summarized in Table 2.

Questionnaires and Interviews

All participants in the planning groups were asked to complete a questionnaire in Chinese immediately after completing the task. The questionnaire consisted of open-ended questions on 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). Additionally, four participants were randomly selected from each group for a retrospective, in-depth interview in Chinese with the researcher. The notes made by the pretask planners and the participants' written responses to the questionnaire served as a basis for the interview questions. The data from the questionnaire and interview were used to help interpret the findings of the statistical analysis.

Table 2. Task conditions

Task condition	*n*	Pretask planning	On-line planning
No planning	14	None	Limited time
Pretask planning	14	10 minutes	Limited time
On-line planning	14	None	Unlimited time

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

Measures

Measures of accuracy, fluency, and complexity were developed to evaluate the quality of the participants' written production. Insofar as possible, these measures were the same as those used in studies of oral production (e.g., Foster & Skehan, 1996; Wendel, 1997; Yuan & Ellis, 2003). However, changes to the fluency measures were needed to make them appropriate to written production as temporal phenomena, such as length of pauses, cannot be measured in writing. Additionally, to determine whether the participants in the OLP group engaged in significantly more on-line planning than the other two groups, the length of time taken and the number of syllables produced by the participants in all three groups were also calculated. In this way, it was possible to determine whether the participants had performed the tasks in accordance with the stipulated planning conditions.

Planning (independent variable) was measured by the following: (a) length of time—the total number of minutes on task was counted for each participant; (b) words—the total number of words produced by each participant; and (c) syllables—the total number of syllables produced by each participant. The length of time and number of words were designed to provide measures of learners' productivity. It was assumed that productivity would vary in accordance with the planning conditions of the three groups and that it would thus be possible to demonstrate the validity of the conditions by reference to these variables.

The dependent variables consisted of the following:

Fluency measures

1. Syllables per minute—the total number of syllables produced divided by the total number of minutes a participant took to complete the task. This is the same measure used by Chenoweth and Hayes (2001). It is preferred to words per minute, which Kellog (1996) recommended as "an average measure of fluency" (p. 65), because it takes the variable lengths of words into account.
2. Number of dysfluencies—the total number of words a participant reformulated (i.e., crossed out and changed) divided by the total number of words produced.

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).[4]
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 or her segments and dividing

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

the total by the total number of segments in the narrative. This procedure was followed to take account of the effect of text length on the type-token ratio.[5]

Accuracy measures

1. Error-free clauses—the percentage of clauses that did not contain any errors. All errors in 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 percentage 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 who coded 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 MSTTR to a low of .82 for dysfluencies, with only two below .90.

Data Analysis

The normal distribution of the three groups' scores on all variables was tested in terms of skewness and kurtosis. A series of one-way ANOVAs were subsequently performed followed by post hoc Scheffé tests where appropriate (i.e., if the *F* score was statistically significant). In the one variable where normal distribution was not evident (i.e., correct verbs), a Kruksal-Wallis Test was run, followed by independent *t*-tests to compare the pairs of groups. The alpha for achieving statistical significance was set at .05. Additionally, effect sizes were calculated using the formula provided in Norris and Ortega (2000, pp. 442–443) to examine the size of the effect of the different kinds of planning on performance of the task. Following Norris and Ortega, effect sizes larger than 0.8 were considered large, sizes between 0.5 and 0.8 medium, between 0.2 and 0.5 small, and less than 0.2 negligible.

RESULTS

Planning: Independent Variables

The means for the independent variables are shown in Table 3. They reveal that the OLP group took longer to complete the task (M = 21.00 min.) than the participants in the NP group (M = 17.00 min.) and those in the PTP group (M = 17.00 min.). The mean lengths of time spent by the NP and PTP groups were identical, corresponding to the time allocated to these groups for completing the writing task. A one-way ANOVA shows that the difference in time taken to complete the task across the groups was statistically significant with the OLP group taking longer than both the NP and PTP groups.

For both syllables and words, there is also an overall statistically significant difference among the three groups. The Scheffé results show that the

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

Table 3. Descriptive statistics and results of ANOVA and Scheffé procedures for independent variables

	Means of planning conditions			ANOVA		Location of significance: Scheffé *p*		
Independent variables	NP	PTP	OLP	*F*	*p*	NP-PTP	NP-OLP	PTP-OLP
Length of time (min.)	17.00	17.00	21.00	47.98*	.000	1.000	.000	.000
Words	181.4	231.1	206.9	7.65*	.002	.002	.146	.177
Syllables	213.14	277.21	242.64	6.74*	.003	.003	.253	.155

*$p < .05$.

PTP group produced significantly more syllables and words than the NP group. The OLP also produced more syllables and words than the NP group, but these differences were not significant. In general, these results indicate that, as anticipated, the OLP group could be distinguished from the PTP and NP groups in terms of the amount of time spent on task. Also, both planning groups could be distinguished from the NP group in terms of the amount of writing produced although these comparisons were only statistically significant in the case of the PTP group.

Dependent Variables

As previously indicated, three aspects of language use were examined to see how the participants in the three treatment conditions performed the writing task. The results of one-way ANOVAs are reported separately for fluency, complexity, and accuracy.[6] Effect sizes are also reported.

Table 4 presents the results for the fluency variables. The PTP group obtained the highest fluency score (M = 16.21 syllables per min.). Thus, the

Table 4. Descriptive statistics, results of ANOVA and Scheffé procedures, and effect sizes for fluency

	M (*SD*) of planning conditions			ANOVA		Location of significance: Scheffé *p* (effect sizes)		
Fluency variables	NP	PTP	OLP	*F*	*p*	NP-PTP	NP-OLP	PTP-OLP
Syllables per minute	12.22 (2.11)	16.21 (3.34)	12.19 (2.70)	9.80*	.000	.002 (1.45)	1.00 (0.01)	.002 (1.33)
Dysfluencies	8.24 (5.73)	4.72 (1.81)	4.86 (2.88)	3.74*	.032	.066 (0.93)	.080 (0.78)	.996 (0.06)

*$p < .05$.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

pretask planners wrote faster than both the no-planners and on-line planners, and these differences were statistically significant ($p < .01$) and reflected in large effect sizes ($d = 1.45$ and 1.33, respectively). The OLP and NP groups were equally fluent with regard to syllables produced per minute. A different picture emerges, however, with regard to the rate of dysfluencies produced by the three groups, with both planning groups manifesting a lower rate than the NP group. Although the differences among the groups were statistically significant overall ($p < .05$), no paired comparison reached statistical significance. However, in both cases, the *d* scores indicated a substantial effect for both types of planning (0.93 in the case of PTP and 0.78 in the case of OLP). Overall, these results indicate that the PTP group wrote faster than the NP and the OLP groups and that the PTP and OLP groups repeated and reformulated less than the NP group.

Three variables were assessed to measure the complexity of language use in the participants' written narratives: syntactic complexity, syntactic variety, and lexical variety. Results are shown in Table 5.

In the case of syntactic complexity, the two planning groups outperformed the NP group. Although these differences did not reach statistical significance, the *d* scores (0.73 and 0.56) indicated a medium level effect for planning. There was little difference in the syntactic complexity of the language produced by the two planning groups ($d = 0.21$). The same ranking for the three groups can be observed for syntactic variety, and in this case the differences were statistically significant. The principal difference lies between the PTP and the OLP and NP. The difference in scores between the PTP and the NP was statistically significant ($p < .01$), and the effect size was large ($d = 4.42$). The difference between the PTP and the OLP approached significance ($p = .079$) with the effect size again large ($d = 0.94$). The OLP also produced more syntactically varied language than the NP ($d = 0.73$), but the difference

Table 5. Descriptive statistics, results of ANOVA and Scheffé procedures, and effect sizes for complexity

	M (*SD*) of planning conditions			ANOVA		Location of significance: Scheffé *p* (effect sizes)		
Complexity variables	NP	PTP	OLP	*F*	*p*	NP-PTP	NP-OLP	PTP-OLP
Syntactic complexity	1.68 (0.26)	2.01 (0.34)	1.92 (0.52)	2.71	.079	— (0.73)	— (0.56)	— (0.21)
Syntactic variety	16.21 (4.01)	22.07 (3.58)	18.86 (3.21)	9.05*	.001	.001 (4.42)	.173 (0.73)	.079 (0.94)
Lexical variety	0.87 (.03)	.88 (.03)	.88 (.04)	0.18	.837	— (0.06)	— (0.05)	— (0.0)

Note. Dashes indicate the Scheffé procedure was not performed.
*$p < .05$.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

was not statistically significant. The three groups performed almost identically where lexical variety (measured by MSTTR) was concerned.

Overall, these results show that both types of planning result in language that is more complex and varied syntactically but that they have no effect on lexical variety. PTP has the larger effect, especially in the case of syntactic variety.

Accuracy was measured in two ways: error-free clauses and error-free verb forms (see Table 6 for the results). The OLP group had the highest mean on both measures, followed by the PTP group, and the NP group had the lowest mean. The ANOVA failed to show these differences to be statistically significant in the case of both variables. However, the difference between groups on error-free clauses did approach significance ($p = .059$), with OLP outscoring the NP ($p = .062$). The d scores indicate a large effect size for on-line planning both when compared to NP ($d = 2.25$) and to PTP ($d = 1.25$). As the scores for correct verbs were not normally distributed, a Kruksal-Wallis Test was also run. This proved significant, $\chi^2 = 6.72$, $p < .05$. Independent t-tests showed that only the difference between the OLP and NP scores was statistically significant, $t = 2.71$, $p < .05$. The effect size for this comparison was large ($d = 1.75$). Medium size effects were also evident in the differences between the OLP and PTP groups ($d = 0.40$) and between the NP and PTP groups ($d = 0.43$). Overall, these results indicate that the opportunity for on-line planning resulted in more accurate written production but that pretask planning had much less effect.

Questionnaires and Interviews

The data obtained from the questionnaires and interviews provide information regarding the nature of the pretask planning activities engaged in by the PTP group and the approach adopted by each group during the writing task.

Table 6. Descriptive statistics, results of ANOVA and Scheffé procedures, and effect sizes for accuracy

	M (*SD*) of planning conditions			ANOVA		Location of significance: Scheffé *p* (effect sizes)		
Accuracy variables	NP	PTP	OLP	*F*	*p*	NP-PTP	NP-OLP	PTP-OLP
Error-free clauses	0.77 (0.01)	0.81 (0.01)	0.86 (0.07)	3.39	.059	.656 (0.0)	.062 (2.25)	.326 (1.25)
Correct verbs	0.85 (0.01)	0.88 (0.13)	0.92 (0.06)	1.98	.152	— (0.43)	— (1.75)	— (0.40)

Note. Dashes indicate the Scheffé procedure was not performed.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

The data were analyzed by counting the number of learners in each group who referred to each of four aspects of the narrative task: (a) understanding the pictures, (b) planning the organization of the story, (c) planning the content of the story, and (d) language planning. Additionally, the number of learners who reported using their L1 was recorded. Table 7 reports the results of this analysis. A minority of participants in each group reported first trying to understand the pictures. The majority of the participants in all three groups reported that they engaged in rhetorical planning of the story (e.g., outlining the key events and establishing links between each picture). Most of the participants also reported working on content (e.g., seeking to be clear, trying to make the story interesting, and adding details) and also attending to the language needed to encode the story. Interestingly, however, they varied considerably in the aspect of language they were planning, with some prioritizing vocabulary and others grammar. A few participants in all three groups specifically reported not engaging in language planning. Very few participants commented that they used their L1.

The PTP group was able to divide their attention to these different aspects between the pretask planning period and the time allocated for writing. In general, they prioritized understanding the pictures, organizing the story, and planning the content during the 10 minutes allocated to pretask planning while they spent their writing time attending to form. In contrast, the OLP and NP groups were obliged to attend to all four aspects of the task while producing the written narrative.

None of the participants mentioned feeling any anxiety when they performed the written task. However, a number of learners in the PTP and NP groups indicated they felt under pressure to finish, and this created a problem for them. For example, one PTP learner commented that her vocabulary was restricted because the limited time made it difficult for her to find appropriate words.

Table 7. Numbers of participants in each group attending to different aspects of writing task

Group	Picture comprehension	Rhetorical planning	Content planning	Language planning	Use of L1
PTP					
Before task	5	12	12	4	0
During task	0	2	4	13	0
OLP					
During task	4	8	11	10	2
NP					
During task	5	14	10	8	1

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

DISCUSSION

We first consider the results for the independent variables and then consider the three research questions. We also examine differences between the results for the effects of planning on written narratives with the parallel results obtained for oral narratives, as reported by Yuan and Ellis (2003). To facilitate this discussion the results of Yuan and Ellis's study and of the present study are summarized in Table 8.

To establish that the two planning conditions worked as anticipated, the length of time the participants in the three groups spent on task and the number of words and syllables they produced were measured. In accordance with the instructions given to the three groups, the OLP group spent longer on task than the other two groups. The fact that the PTP group produced more words and syllables than the participants in the NP group indicates that the pretask planners did in fact use the time to plan their written narratives. This conclusion is also supported by the participants' answers to the posttask questionnaire. Overall, the three groups appear to have performed in accordance with the instructions they received and with the researchers' expectations.

The first research question addressed the effects of pretask planning on fluency, complexity, and accuracy in written narratives. The PTP group outperformed the NP group in the number of syllables per minute. Thus, providing an opportunity for pretask planning was not only advantageous in terms of the quantity of writing produced but also in terms of the production speed. It should be noted, however, that, although the pretask planners wrote more quickly, they spent more time on task than the on-line planners overall—a total of 27 minutes (10 minutes for planning followed by 17 minutes for writ-

Table 8. Summary of the effects of two types of planning on L2 oral and written narratives

Task medium	Planning type	Fluency	Complexity	Accuracy
Oral	Pretask	No effect	Increased syntactic complexity	No effect
	On-line	No effect	Increased syntactic complexity	Increased accuracy
Written	Pretask	Increased fluency and fewer dysfluencies	Marked increase in syntactic complexity and variety	Little effect
	On-line	Decreased fluency but fewer dysfluencies (tendency)	Some increase in syntactic complexity and variety but not statistically significant	Increased accuracy

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

ing), as opposed to 17 minutes for the no-planners. Pretask planning also resulted in fewer dysfluencies in the textual output, a difference that approached statistical significance. The pretask planners produced fewer false starts and self-corrections.

In terms of the model of writing in Figure 1, it can be surmised that pretask planning aids fluency in writing in two principal ways. First, it facilitates process and text planning for content and organization. This is reflected in the pretask planners' responses to the questionnaire and their comments in the posttask interview. Twelve of the 14 pretask planners reported engaging in rhetorical and content planning. A writer who has a clear idea as to the text type required (narrative) and the point of the story depicted in the pictures, organizes the information that needs to be conveyed, establishes the setting and describes the characters, identifies the main events, and evaluates them will find the pressure on working memory lessened during on-line assembly (Raab, 1992, cited by Zimmerman, 2000). Second, pretask planning may help to increase L2 writers' confidence in their ability to write clearly and effectively and, for this affective reason, may reduce their need to engage in extensive monitoring. Zimmerman found that writers revise more when writing in their L2 than in their L1; thus, one of the effects of allowing time for pretask planning may be to reduce the number of revisions undertaken in L2 writing, with the result that it resembles L1 writing more closely. Chenoweth and Hayes (2001) found that L2 writers who were more proficient wrote more fluently than less proficient writers; pretask planning, therefore, may compensate for lack of L2 proficiency where fluency is concerned.

Pretask planning also has some effect on complexity. The pretask planners used more subordination and a greater variety of verb forms in their written narratives than the NP group although only the comparison for syntactic variety achieved statistical significance. Pretask planning had no effect on lexical variety. These results suggest that pretask planning time was used by the learners to focus attention on the propositional content of the story by specifying the key actions depicted in the pictures (i.e., identifying the main narrative events). This focus on planning the events was presumably accompanied by a search for verb forms to encode temporal and modal meanings. In this respect, it can be noted that, in the posttask interviews, a number of the PTP-group participants specifically commented on having attended to verb tense when planning the task. These verb forms were then stored and subsequently accessed during the translating process. That pretask planning has no effect on lexical variety may reflect the fact that, given the importance of locating the relevant vocabulary to encode the propositional content of the story, all the writers prioritized lexical search during on-line assembly. In Levelt's (1989) speaking model, lexis takes precedence over grammar during formulation (or translation in Kellog's, 1996, model). Given that writing, even when pressured, allows more time for formulation, this is even more likely to be the case.

Pretask planning had some effect on linguistic accuracy, but the comparisons for both error-free clauses and correct verbs did not achieve statistical

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

significance. Enhanced accuracy in writing may be due primarily to the monitoring that occurs when writers revise the output of translation (as in Kellog's model), using explicit knowledge of their L1, L2, or both. Pretask planning does not contribute to this in any significant way. The variable results obtained for the effects of pretask planning on accuracy in previous task-based research may simply reflect the extent to which the participants in those studies had sufficient time to engage in editing during the task, a variable that was not controlled for.

In summary, pretask planning has a marked effect on written fluency, a relatively strong effect on linguistic complexity (especially the variety of verb forms), and very little effect on accuracy. We have suggested that this can be best explained by hypothesizing that pretask planning assists internal goal setting, the rhetorical organization of the text to be produced, and the propositional content to be encoded. This reduces the pressure on the central executive in working memory and thus facilitates the process of translating what has been planned into verbal schema, even when this has to be undertaken under pressure of limited time. The opportunity for pretask planning may also add to the learners' confidence during task performance. In the posttask interviews, the pretask planners in particular reported that they felt they were able to handle the task easily. However, pretask planning does not contribute to editing when writers are pressured to write quickly.

These results are comparable to the results obtained for pretask planning of oral narratives in a number of ways. Yuan and Ellis (2003) found that pretask planning resulted in more speech and greater complexity in comparison to no planning. It had no statistically significant effect on fluency or on accuracy.[7] Thus, pretask planning would appear to increase quantity of output and enhance some aspects of linguistic complexity but not accuracy, in both speech and writing. The main difference arises with regard to fluency. However, other studies (e.g., Crookes, 1989; Foster & Skehan, 1996) have shown a positive effect for pretask planning on fluency in oral performance. It is possible that Yuan and Ellis failed to find an effect in their study because the pretask planners were only allowed limited time for on-line assembly, whereas in the other studies on-line production was not pressured. The lack of time for on-line assembly may have induced anxiety (that the participants performing the oral task reported experiencing), which in turn had a negative effect on oral fluency. In this study, the pretask planners were also given limited time to perform the task, but perhaps this was less threatening, given that writing, even when pressured, does not require learners to perform in real-time.

The second research question concerned the effects of on-line planning. As might be expected, encouraging learners to plan on-line does not result in greater fluency. However, it does not appear to inhibit fluency, either, given that the words per minute score for the OLP group was almost identical to that of the NP group, which was pressured to write quickly. Furthermore, it tends to reduce dysfluencies (although the difference between the OLP and the NP did not reach statistical significance, the effect size was quite large).

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

Writers may take advantage of the time available for on-line planning to monitor their internally processed output during translation before they execute the text. In contrast, when pressured to write quickly, writers rely more on editing and correcting actual textual output. On-line planning has some effect on complexity, but the differences in both syntactic complexity and variety were not statistically significant. It had a clear effect on accuracy, with the OLP group producing text that was more accurate in terms of both error-free clauses and correct verbs than the NP group and with the difference achieving statistical significance in the case of the former variable. It is likely that the on-line planners used the additional time at their disposal to attend carefully to linguistic accuracy by editing their internal and external output. Writing allows learners access to observable units of text and thereby induces attention to form. When on-line assembly is unpressured, this inherent tendency in writing may be accentuated. In support of this, it can be remarked that the on-line planners noted in their posttask interviews that they attempted to attend to everything, which suggests that they monitored the output of translation extensively before and after execution.

Yuan and Ellis (2003) found that opportunities for extensive on-line planning had no effect on fluency but promoted syntactic complexity and accuracy in comparison to no planning. It also resulted in greater lexical variety than pretask planning. The main difference between the results for the oral and written tasks, then, concerns complexity. It should be noted, however, that the means of the OLP group in the present study for syntactic complexity ($M = 1.92$) and syntactic variety ($M = 18.86$) were higher than those for the NP group ($M = 1.68$ and 16.21, respectively) and in this respect mirror the results obtained for the oral narrative in Yuan and Ellis. That the differences were not statistically significant may reflect the fact that, given the nature of writing, the NP group found time to reflect on propositional content despite the pressure to write rapidly. Similarly, the resemblance in lexical variety scores in the three groups may indicate that in the written task all learners had sufficient time for lexical searching and prioritized this aspect of verbal processing.

The third research question addressed the relative effects of the two types of planning. The pretask planners were found to be more fluent than the on-line planners and tended to use a greater variety of verb forms ($d = 0.94$). They were less accurate than the on-line planners although this difference was not statistically significant. The results of this study, therefore, suggest that the two types of planning have different effects on the quality of textual output when measured by means of specific linguistic measures. We explain this by positing that they facilitate different components of the writing process. That is, pretask planning assists Kellog's (1996) formulation component, whereas unpressured on-line planning allows time for monitoring.

The results obtained in the present study differ from those in Yuan and Ellis (2003) for complexity. Whereas this study found some evidence of greater complexity in the PTP group, Yuan and Ellis found no difference between the

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

two groups. We hypothesize that this again reflects the fact that, in comparison to the oral task in Yuan and Ellis, the writing task did provide the pretask planners with opportunities for some controlled on-line planning, despite the time pressure. Thus, the pretask planners in this study benefited from time for both planning and translating.

In summary, it is clear that pretask planning enhances learner output in a written task. This is manifested in greater quantity, fluency, and complexity of language, although such planning appears to have little effect on accuracy. It should be noted, however, that the pretask planners' total time on task exceeded that of both the no-planners and the on-line planners. Thus, as Hayes and Gradwohl Nash (1996) pointed out, it cannot be concluded that pretask planning is more efficient than on-line planning. The opportunity to plan on-line also confers benefits, but of a different kind. Although it does not lead to greater fluency (as might be expected), it does reduce the number of dysfluencies. It also results in more accurate use of the L2. Arguably, if the goal is to ensure that L2 writers produce their highest quality work, they need time for both types of planning.

CONCLUSION

The purpose of this article was to explore the effects of two types of planning (pretask and on-line) on textual output. We did not attempt to explore the actual strategies and processes involved in these two types of planning in any detail; rather, we used Kellog's (1996) model of L2 writing (as in Figure 1) to explain the effects we found. We believe that our results are explicable in terms of this model. Table 9 identifies the components of the writing process likely to have been attended to by writers in the different planning conditions. Clearly,

Table 9. Attention to different components of the composing process according to planning condition

	Composing process components		
Planning condition	Formulation	Execution	Monitoring
PTP +Pretask planning time −On-line planning time	Extensive	Yes	Little
OLP −Pretask planning time +On-line planning time	Some	Yes	Extensive
NP −Pretask planning time −On-line planning time	Limited	Yes	Little

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

however, there is a need to examine more closely the relationship between "process" and "product" using the same experimental design but probing more deeply what L2 writers actually do when they engage in pretask and on-line planning. Ideally, the pretask and on-line planners need to spend the same amount of total time on task although this might be difficult to achieve without imposing undue pressure on the pretask planners when they begin to write. There is also a need to include a third experimental condition, in which the participants have the opportunity for both pretask and unpressured on-line planning.

Although there have been studies investigating what L1 writers do when they plan, few have focused on the effects of planning on L2 written output. This is surprising, given the importance attached to the "planning stage" in pedagogic discussions of process writing. Hedge (2000), for example, unequivocally stated, "good writers concentrate on the overall meaning and organization of a text, and engage in planning activities" (p. 305). This study, then, begins to show that planning does impact significantly on the quantity and quality of L2 writing. Furthermore, it suggests that pretask and on-line planning have somewhat different effects. This has important implications for both writing pedagogy and testing. That is, teachers may be able to manipulate the aspects of writing (fluency, complexity, and accuracy) that L2 writers attend to by varying the task conditions to allow sometimes for pretask planning, sometimes for unpressured on-line planning and sometimes for both. Finally, testers who wish to enable L2 writers to present their best products for assessment may need to ensure that opportunities for both types of planning are available to testees.

(*Received 15 June 2003*)

NOTES

1. The stressful nature of speaking is well documented (see, e.g., Horwitz, Horwitz, & Cope, 1986). Writing can also be anxiety inducing (see Cheng, Horwitz, & Schallert, 1999) but, we would maintain, usually less so than speaking—a view supported by the comments made by the participants in the posttask interviews.

2. The Higher Education Bureau Examination is a compulsory English examination for all students who wish to major in English at a Chinese university. It consists of simple questions and answers, a spoken discourse on an unseen topic, and a story retelling. The participants in this study all scored between a B+ and an A on this test, with most obtaining an A−.

3. One anonymous *SSLA* reviewer pointed out that the pretest material should ideally have consisted of a writing task similar to that used in the study and performed without time allowed for either pretask and on-line planning. We acknowledge that this would have provided a basis for demonstrating the groups' equivalence in fluency, complexity, and accuracy. However, we argue that, because the groups were randomly constituted and because we took steps to ensure there were no significant group differences in either the total TOEFL or TOEFL listening scores, we have sufficient grounds to claim that the groups were equivalent in general English proficiency and thus likely to be also equivalent with regard to the fluency, complexity, and accuracy of their written production.

4. A c-unit is defined as an utterance that consists of a single complete sentence, phrase, or word and that has a clear semantic and pragmatic meaning in the context in which it occurs. In effect, it is the same as a T-unit except that it includes elliptical utterances.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

5. Malvern and Richards (2002) pointed out that MSTTR "removes the problem of variation in sample size and wastes less data than if all analyses were performed on a standard number of words" (p. 88). However, they also acknowledged problems with it—for example, "very short segments (even those of 100 tokens) are likely to distort results because they are not sensitive to repetition of words beyond the boundary of their own segment" (p. 88). In our study, written texts were divided into segments of 40 words. Malvern and Richards proposed a new measure based on D-values. However, we lacked the software needed to compute this and following advice from Richards (personal communication) relied on MSTTR. We acknowledge the limitations of this measure but argue it is a more valid measure than standard type-token ratio.

6. One anonymous *SSLA* reviewer suggested that, given the range of dependent variables, a MANOVA should have been performed. However, the effects of the treatments were examined only by three one-way ANOVAs, minimizing the risk of a Type 1 error. Further, Keselman et al. (1998) argued that there is very limited empirical support for a MANOVA univariate data analysis strategy.

7. It should be noted, however, that the PTP group in Yuan and Ellis's (2003) study performed more fluently than either the NP or the OLP groups, and when compared with the OLP group, the difference was statistically significant. Thus, there is some evidence that pretask planning can result in greater oral fluency.

REFERENCES

Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Mahwah, NJ: Erlbaum.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology*, *29*, 591–620.

Brown, J., McDonald, J., Brown, T., & Carr, T. (1988). Adapting to process demands in discourse production: The case of handwriting. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 45–59.

Cheng, Y., Horwitz, E., & Schallert, D. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, *49*, 417–446.

Chenoweth, A., & Hayes, J. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, *18*, 80–98.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, *11*, 367–383.

Cumming, A. (in press). Models of second-language composing processes. In R. Weissberg (Ed.), *ESL writers in the college composition classroom: Challenges and opportunities*. Urbana, IL: National Council of Teachers of English.

De Larios, J., Marin, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, *51*, 497–538.

Educational Testing Service. (1987). *Reading for TOEFL Workbook*. Princeton, NJ: Author.

Elbow, P. (1981). *Writing with power*. Oxford: Oxford University Press.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style in the use of the past tense. *Studies in Second Language Acquisition*, *9*, 12–20.

Flower, L., & Hayes, J. (1980). The dynamics of composing: Making plans and juggling constraints. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Mahwah, NJ: Erlbaum.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*, 299–323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354–375.

Friedlander, A. (1990). Composing in English: Effects of a first language on writing in English as a second language. In B. Kroll (Ed.), *Second language writing* (pp. 109–125). New York: Cambridge University Press.

Grabe, W. (2001). Notes on a theory of second language writing. In T. Silva & P. Matsuda (Eds.), *On second language writing* (pp. 39–57). Mahwah, NJ: Erlbaum.

Grabe, W., & Kaplan, W. (1996). *Theory and practice of writing: An applied linguistic perspective*. London: Longman.

Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 1–27). Mahwah, NJ: Erlbaum.

Hayes, J., & Gradwohl Nash, J. (1996). On the nature of planning in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 29–55). Mahwah, NJ: Erlbaum.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press

Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford University Press.
Heaton, J. (1975). *Beginning composition through pictures*. London: Longman.
Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*, 125–132.
Hulstijn, J., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning*, *34*, 23–43.
Kellog, R. (1996). A model of working memory in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 57–71). Mahwah, NJ: Erlbaum.
Keselman, H., Huberty, C., Lix, L., Olejnik, S., Cribbie, R., Donahue, B., et al. (1998). Statistical practices of educational researchers: An analysis of the ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, *68*, 350–386.
Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second language writing* (pp. 140–154). New York: Cambridge University Press.
Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*, 85–104.
Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*, 52–83.
Norris, J., & Ortega L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, *21*, 108–148.
Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, *45*, 99–140.
Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*, 185–211.
Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*, 93–120.
Wendel, J. (1997). *Planning and second language narrative production*. Unpublished doctoral dissertation, Temple University Japan, Tokyo.
Whalen, J., & Menard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning*, *45*, 381–418.
Woodhall, B. (2002). Language switching: Using the first language while writing in a second language. *Journal of Second Language Writing*, *11*, 7–28.
Yuan, F., & Ellis, R. (2003). The effects of pretask planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*, 1–27.
Zimmerman, R. (2000). L2 writing: Subprocesses, a model of formulating and empirical findings. *Learning and Instruction*, *10*, 73–99.

https://doi.org/10.1017/S0272263104026130 Published online by Cambridge University Press