

THE INFLUENCE OF PLANNING AND TASK TYPE ON SECOND LANGUAGE PERFORMANCE

Pauline Foster and Peter Skehan

Thames Valley University

This study focuses on the impact of different variables on the nature of language performance in the context of task-based instruction. Characteristics of tasks are discussed, and then a framework is offered that can organize the nature of task-based instruction and relevant research. The framework is used to generate predictions regarding the effects of three different tasks (Personal Information Exchange, Narrative, and Decision-Making) and three different implementation conditions for each task (unplanned, planned but without detail, detailed planning) on the variables of fluency, complexity, and accuracy. The study reports strong effects of planning on fluency and clear effects also on complexity, with a linear relationship between degree of planning and degree of complexity. However, a more complex relationship was discovered between planning and accuracy, with the most accurate performance produced by the less detailed planners. In addition, interactions were found between task type and planning conditions, such that the effects of planning were greater with the Narrative and Decision-Making tasks than with the Personal Information Exchange task. The results are discussed in terms of an attentional model of learning and performance and highlight the importance of tradeoff effects between the goals of complexity and accuracy in the context of the use of limited capacity attentional resources. The study contributes to the development of cognitive models of second language performance and addresses a number of pedagogic issues.

The authors thank four anonymous *SSLA* reviewers whose comments on an earlier draft of this article were very helpful. We also thank the staff and students at Richmond Adult and Community College for their considerable cooperation. Without them the research reported here would have been impossible. Finally, thanks are also due to the Centre for Applied Linguistic Research at Thames Valley University and to the Economic and Social Research Council, whose financial support was crucial in enabling the research to proceed.

Recent work applying concepts from cognitive psychology to language teaching have suggested that (a) attention and noticing are central for second language development (Schmidt, 1990, 1994) and (b) attentional resources are limited (VanPatten, 1990). Decisions have to be made about their allocation (Anderson, 1995), and decisions to attend to one area are usually made at the expense of attending to another. In the context of second language learning, the main competition for attentional resources is between meaning and form, with first priority generally being given to meaning (VanPatten, 1990).

These developments have important implications for language teaching. In recent years there has been a significant move from controlled, structured approaches to syllabus and teaching to the use of more communicative materials meant to promote interaction (Harmer, 1991). This, in turn, is seen as catalyzing acquisitional processes. One question that has arisen from this change is whether interaction should be viewed as an adjunct of traditional approaches, a more engaging version of conventional production activities (Rivers, 1981), or interaction should actually constitute the language learning experience in a more self-sufficient way. In this respect, recent proposals for task-based approaches (Long & Crookes, 1991, 1993; Prabhu, 1987) represent a clear case of the more extreme interaction-based position, as they advocate tasks as the *unit* of syllabus design and teaching. Tasks are defined as activities that are meaning-focused and outcome-evaluated and have some sort of real-world relationship. In other words, transacting tasks will push forward interlanguage development because the demands that tasks make will engage the very processes that lead to acquisition.

This approach to syllabus design and language pedagogy requires us to know more about the general area of task-based instruction. To try to achieve this, this article will explore two relevant areas: task analysis and task implementation. Recent proposals regarding task analysis have divided generally into (a) those emphasizing the potential of different tasks to generate patterns of language use that are deemed “better” and (b) those concerned with establishing task difficulty (Long, 1989). In the former category, Duff (1986) proposes a distinction between *convergent* (one answer must be agreed, e.g., on how to allocate blame in a moral problem) and *divergent* (disagreement, e.g., in a discussion task, is regarded as natural) tasks. She proposes that convergent tasks have a greater potential to generate negotiation of meaning, a theoretically valued form of conversational interaction. Long (1989) distinguishes between *open* and *closed* tasks, again relating the distinction to the extent to which tasks lead to greater use of conversational gambits such as confirmation checks, clarification requests, and other features that are seen as implicated in interlanguage change. Similarly, Pica, Kanagy, and Falodun (1993) propose that tasks can be analyzed in terms of the following scheme:

Interactional activity

Interactional relationship, that is, information held by different participants

Interactional requirements, that is, required or optional participation

Communicational goal

Goal orientation: convergent or divergent

Outcome option: one solution only, or more than one

This scheme enables tasks to be chosen that yield a “better” quality of language for acquisition, assuming that the negotiation of meaning, which would be promoted, does indeed reflect a more malleable interlanguage system and greater scaffolding and feedback at particularly opportune moments. Aston (1986) argues that such negotiation is more irritating to learners than beneficial, and Foster (1994) proposes that relatively few subjects in research studies may account for virtually all the negotiation of meaning that is found; that is, most learners do not negotiate at all!

The alternative perspective on task analysis is to propose features that enable tasks to be ranked in difficulty. Brown, Anderson, Shilcock, and Yule (1984), based on extensive empirical work, propose two dimensions for this, one concerned with type of information, the other with scale. The former refers to the increasing difficulty of tasks as they move from static to dynamic to abstract, whereas scale concerns the number of participants or objects contained in the task as well as the nature of the relationships between these elements. Even so, the tasks on which this research was based are restricted in type. Candlin (1987) proposes a wider ranging (but speculative) scheme for assessing task difficulty, and for incorporating cognitive load, communicative stress, code complexity, clarity of task goals, and familiarity of task type. Skehan (1996) extends and slightly reorganizes this scheme into the following categories:

Language factors

Syntactic complexity and range

Lexical complexity and range

Redundancy and variety

Cognitive factors

Familiarity

Familiarity of material in the task

Familiarity of task-type and discourse genre

On-line processing

Reasoning operations required

Nature of input material used in the task

Degree of organization of input material

In this way, a clear distinction is made between language and cognition, as well as the consequences for task difficulty that arise from each of these. Then, within

cognitive factors, a further distinction can be made. On the one hand there is accessing knowledge sources directly, that is, drawing upon knowledge already held in memory that is relevant to completing a particular task. On the other hand there is needing to work out the answer to a task; that is, an active intellectual engagement with the task is called for because understanding the material involved, or arriving at a solution, cannot simply be based on accessing existing knowledge and using it untransformed. This scheme, too, is speculative, but it is an important motivating influence upon the present study.

Regarding task implementation, a number of studies have suggested ways in which pedagogic decisions influence performance. Ellis (1987), in a picture-story narration task, investigated the effects of engaging what he termed different degrees of planned discourse. Subjects in the most planned condition were required to write a story to a set of (coherently structured) picture prompts. A second condition required these subjects to tell the same story orally, but without access to their previous written versions. Finally, the least planned condition was one in which subjects had to tell a story orally to a new set of picture prompts. Ellis (1987) found that accuracy on the regular (rule-governed) form of the past tense declined according to lack of planned discourse being engaged, whereas performance on the irregular (lexical) past form was hardly affected by these conditions. In contrast, Crookes (1989) reported a study in which providing subjects with pretask planning time leads to greater complexity and variety of language but no significant increases in accuracy. In this respect, Crookes proposed that the Ellis (1987) study confounded modality and planning time, rendering its results suspect. However, a problem with the Crookes study is that it is difficult to know what subjects actually did during the planning time available to them. It is noteworthy, in this regard, that the standard deviations for the measures reported tended to be higher for the planning groups for the complexity measures, suggesting that some subjects interpreted the planning instructions in different ways, using the time to prepare differently. Research attempting to influence the nature of the planning that takes place would therefore seem desirable.

Other manipulations of task conditions are also possible. Harrison (1986), for example, suggested that for testing as well as teaching it would be worth investigating the impact on performance of introducing surprise elements midtask, to explore how well different learners cope with this condition, on the assumption that it is likely to be disruptive to some degree but nonetheless typical of everyday language use. Conversely, it may be useful to explore how supporting learners with, for example, visual representation of the information required in tasks may ease processing demands and improve performance (Loschky & Bley-Vroman, 1993; D. Willis, 1990). It is also possible to introduce posttask manipulation. Here the central issue is that if one is using a task-based approach one does not want to interfere with performance midtask, because to do so would compromise one of the central defining qualities of tasks: they are meaning-focused. But a focus on form (rather than a focus on forms, i.e., form in general rather than any form in particular) is very important if language development is to occur (Doughty, 1991) because such a focus makes noticing more likely and also enables language teaching goals of accuracy

Table 1. A framework for task implementation

| Stage | Goal | Typical Techniques |
|------------------|---|--|
| Pretask | | |
| Linguistic | Introduce new forms to interlanguage repertoire | Explicit and implicit teaching |
| Cognitive | Reduce cognitive load | Consciousness-raising |
| | | Plan linguistically and cognitively |
| | | Observe similar tasks |
| | Push learners to express more complex ideas | Plan |
| | | Observe |
| Midtask | | |
| Task choice | Balance difficulty of task | Use analytic scheme |
| Task calibration | Increase or reduce difficulty | Introduce surprise |
| | | Provide (visual) support |
| Posttask | Raise consciousness for a focus on form | Use public performance and posttask activities |

and complexity (Skehan, 1992) to be pursued more effectively. To that end, it has been proposed (J. Willis & Willis, 1988) that posttask activities with a strong focus on form can have a beneficial effect on (earlier) task performance itself, provided that learners themselves make the connection between task completion and subsequent posttask activity. In this way, posttask activities, if salient in learners' attention, can infiltrate a focus on form into the earlier task completion and reduce the likelihood that the learners engaged in such task completion will allocate attention exclusively to meaning.

Based on studies such as these, Skehan (1996) proposes a framework to organize methods of task implementation that allow principled decisions to be made regarding the attentional focus and pedagogic goals of different activities. The framework is shown in Table 1.

The framework draws upon existing research but then elaborates a more comprehensive perspective on how task-based approaches can be implemented. Obviously, the components of the framework constitute claims as to variables that will be important in task-based studies, and so underpinning research needs to be carried out to support or disconfirm the claims implicit in Table 1. We will report on research here that bears upon the role of planning as an aspect of pretask activities. Prior to that, however, we need to discuss alternative methods of *assessing* task-based performance and the relationship these different measures have to underlying theory.

In considering performance on tasks and following researchers such as Crookes (1989), Skehan (1992, 1996) makes a three-way distinction between complexity, accuracy, and fluency. The first two of these concern form but with a significant difference in emphasis. Complexity emphasizes the organization of what is said and draws attention to the progressively more elaborate language that may be used, as well as a greater variety of syntactic patterning. Complexity is likely to reflect a willingness, on the learner's part, to engage in restructuring as more complex subsys-

tems of language are developed. It is also 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.

The other aspect of form is accuracy. In this case the focus is on freedom from error, based on whatever language is used. In other words, a measure of accuracy may not reflect complexity of language and may be the result of relatively simple, well-controlled forms being used to achieve a more target-like use of language. In a sense, therefore, accuracy, as a reflection of a focus on form, may have a more conservative orientation, whereas complexity may capture a greater willingness to experiment and to take risks. Complexity, in other words, connects with change and the opportunities for development and growth in the interlanguage system. Accuracy, in contrast, concerns control at a particular interlanguage level.

Finally, we need to consider the role of fluency. Skehan (1992, 1996), following Schmidt (1990), takes fluency to reflect the primacy of meaning and the capacity to cope with real-time communication. It may therefore prioritize lexicalized language (Ellis, 1987) and the way learners, during tasks, avoid rule-based, constructed language, instead preferring to use more idiom-based language (Sinclair, 1991) to enable communication to proceed more smoothly. Fluency also reflects the effectiveness of the planning process and the way propositions can be orchestrated into effective, ongoing discourse.

It is also important to consider how each of these variables can be operationalized. A first question concerns whether to use general or specific measures. Ellis (1987) used specific measures of accuracy, such as measuring error rates with three past-tense forms. Crookes (1989), similarly, focused on use of articles and third-person -s. Robinson, Ting, and Urwin (1995), in contrast, reported the use of more generalized measures for fluency (counting the number of pauses and the number of words per utterance) and complexity (calculating the number of embeddings per S-node, as well as lexical density). On occasion, clearly, specific measures are appropriate, as in the Ellis (1987) study in which the research design generated large numbers of tokens of the past tense and in which the past tense was central to the task. But when tasks do not provoke hypotheses about the use of specific forms, it would seem more appropriate to use generalized measures. Above all, such measures are likely to be more sensitive to differences in experimental conditions, detecting weaker effects because these are likely to be based on more variance. Crookes (1989) relied principally on specific measures (target-like use of the article system and of plural -s), and this may have been a factor in the lack of significance reported. Given the exploratory state of planning research, a general approach to measuring accuracy would appear more defensible at present.

But this still leaves the question of exactly which general measures should best be used. The general principle proposed is that useful variance should be maximized. To that end, as a measure of accuracy, the calculation of error-free clauses has merit. The unit involved, the clause, is relatively short, so that there is considerable scope for subjects to get moderately high scores. Clauses would also generalize across a range of different specific forms and not be dependent on any one of them. Finally, such a measure should produce good intersubject variation.

Turning to complexity, more generalized measures have been widely used, as with Robinson et al. (1995) and Crookes (1989), who used the number of subordinate clauses per utterance and per T-unit. To the extent that subordination, as an index, reflects a greater degree of internal structuring of speech, it captures a great deal of useful variance and so is a satisfactory measure. However, it can be argued that measures based on T-units ignore useful data. Hunt's (1966) definition of the T-unit (as a main clause plus embedded or attached subordinate clauses) excluded ellipsis and is therefore often unsatisfactory in dealing with spoken interaction where ellipsis quite naturally abounds. The c-unit, defined by Brock (1986) as an independent utterance providing referential or pragmatic meaning, allows for ellipsis and is thus a more sensitive measure for the spoken language. It is considered more appropriate for this study to capture complexity within the spoken language produced. But in addition to measuring complexity through subordination, one can also try to measure the range of syntactic forms used, as these reflect a capacity to draw upon a greater variety of language. If only base forms are used, such as simple present or simple past tenses, there would seem to be little impetus for interlanguage change to occur through the use of more challenging forms. To explore this possibility, it would seem worthwhile to examine measures of syntactic variety. Following Crookes (1989), we will focus on verbs and examine such areas as tense, aspect, voice, and modality.

Finally, we need to consider measures of conversational fluency. There is no generally accepted operationalization of fluency, although one can distinguish among different aspects. One is the capacity to engage in continued performance. This would suggest measures such as pauses, the total amount of silence, and perhaps repetition and hesitation as appropriate indices. Another aspect of fluency is the need to engage in more frequent repairs of breakdown in speech. This can be measured through indices of replacement, false starts, and reformulation. Finally, given arguments presented earlier, one might assess fluency through the importance of lexicalized communication. This study will not explore the last of these possibilities, useful though it would be, as there are no clear methods for establishing lexicalized units of production. Consequently, the measures used will reflect *fluency as continued performance* and *fluency as repair avoidance*.

Drawing on the preceding review, we can now formulate a number of hypotheses regarding the effects of planning on the nature of performance on language learning tasks. These hypotheses relate to the three aspects of performance already discussed: fluency, complexity, and accuracy. The overriding hypothesis is that planning will have positive effects on performance in each of these areas, but this general claim can be broken down into several more detailed hypotheses:

Hypothesis 1: Under planned conditions, there will be greater fluency in language; that is, there will be fewer repetitions, fewer reformulations, fewer replacements, fewer false starts, fewer pauses, and a smaller silence total. This follows from the preparation that can be accomplished during planning. Ideas can be organized to enable more fluid expression, less on-line attention is needed to handle ongoing discourse processing, and lexicalized language units can have been prepared to ease processing even further.

Hypothesis 2: Under planned conditions, there will be greater complexity of language; that is, there will be a greater number of clauses per c-unit. This relates to the Crookes (1989) study. Planning can enable more ambitious ideas to be attempted and will enable the relationships between ideas to be expressed more clearly.

Hypothesis 3: Under planned conditions, there will be greater variability in language; that is, there will be a wider range in the tense, aspect, modality, and voice of verbs. This also follows from Crookes (1989). Nonplanners will need to engage in more on-line processing and, as a result, will tend to rely on a narrower range of linguistic forms. Planners will be able to access a wider repertoire of linguistic forms, partly because of freer on-line resources and partly because such forms will have been made salient from the planning phase itself.

Hypothesis 4: Under planned conditions, there will be greater accuracy in language; that is, there will be a higher proportion of error-free to total clauses. This relates to the study by Ellis (1987) in which significant differences in accuracy were found and to that by Crookes (1989) in which there was a nonsignificant trend toward accuracy, though only specific accuracy measures were used. It is proposed that under the planning condition greater accuracy will result partly from the greater attention that is available on-line (Tarone, 1985) and partly from the activity in the planning phase itself that is directed toward achieving accuracy.

Hypothesis 5: The effects predicted in Hypotheses 1–4 will all be greater when planning is carried out at a more detailed level. The motivation for this hypothesis relates to the effectiveness with which planning is directed toward task performance and to the larger standard deviations for the planners' complexity scores reported in Crookes (1989). It is proposed that more detailed planning will lead subjects to use planning time more effectively and in ways that aid subsequent retrieval of the products of planning during task performance.

Hypothesis 6: The effects predicted in Hypotheses 1–5 will all be greater for tasks that are more cognitively demanding, that is, those based on less familiar information and requiring more on-line processing. Such tasks provide a greater challenge for subjects, as they require more attentional resources to handle cognitive content and, correspondingly, provide greater scope for planning to prepare for the subsequent task performance.

METHOD

The study is essentially a multifactorial design, examining the effects of task choice and implementation conditions on a range of measures of language performance. We will describe the tasks themselves, the implementation conditions, the subjects, the setting, the overall research design, and then the dependent variables.

Tasks

All tasks were carried out by subjects in dyads. The three tasks used were a personal information exchange, a narrative based on pictures, and a decision-making task. The choice of these types was based on an analysis of tasks commonly used in current English language teaching textbooks. They were hypothesized to require different levels of attention on the part of the subjects, with progressively less familiar and less predictable information causing an increasingly taxing cognitive load and, as a consequence, influencing performance on the task.

The *Personal Information Exchange* task required subjects to describe to the other member of the dyad how to get to his or her home from the college that subjects were attending and then to turn off a gas cooker that had been left on. As it involved accessing information well known to the speaker and possibly already rehearsed in English, it was seen as requiring the least cognitive effort and allowing the greatest attention to language form. Moreover, it was reasoned that the nature of the task would require relatively simple linguistic forms to be used.

For the *Narrative* task, each member of the dyad had to construct a storyline from a set of five pictures that were loosely but not obviously connected and to relay their ideas to each other. This task involved encoding new, visual information into linguistic form and required some degree of imagination. It was seen as giving scope for more complex language but also demanding greater cognitive effort, therefore allowing less attention to be devoted to form.

For the *Decision-Making* task, subjects were asked to act as judges at the trials of a list of offenders and to reach an agreement with their partner on a suitable prison sentence for each. This task involved considering a lot of new information (i.e., the facts of each case), evaluating it, and then defending an opinion against any objections from the other member of the dyad. This task was considered to place the heaviest cognitive load upon the subject and to allow the least attentional resources to be given to language form. At the same time, the process of trying to reach an agreement on a series of difficult questions was considered most likely to require the use of complex language.¹

This characterization suggests that the Personal task should be easiest and most accessible to learners and that the Narrative and Decision tasks are similar to one another in that they provide relatively low familiarity but a high computational load. Looking at these latter two tasks in more detail, one might propose that the Decision task is slightly more familiar than the Narrative task, in that the moral values relevant to each judgment ought not to be totally unknown to the subjects in the study. Conversely, one might argue that the more unpredictable interactions in the Decision task make it more difficult. It should also be the case that the Narrative task is eased through the visual and “plot” support that is involved. On balance, then, the Narrative task is judged to be easier than the Decision task.

Planning Condition

Three groups of students were involved in the study: one control and two experimental. The control group dyads did the tasks with only a brief introduction to ensure that they all understood what was required of them. The experimental dyads had the same brief introduction, followed by 10 minutes of individual planning time (as in Crookes, 1989). To maximize the chances that they would indeed engage in planning, they were asked to make notes about what they were going to say but were told that these notes would be taken away before they began to speak. The planning condition was operationalized as detailed and undetailed; that is, half of the experimental dyads received guidance on how they might use the 10 minutes to consider the syntax, lexis, content, and organization of what they would say, and

the other half received no guidance and were simply told to plan. In this way, it was possible to investigate any effect that quality of planning might have on task performance.

Subjects and Setting

The subjects were 32 preintermediate-level students studying English as a foreign language 6 hours per week at a local college. They came from a wide variety of L1 backgrounds and were all between 18 and 30 years old, and all but three were female. They had been placed in one of four preintermediate classes on the basis of a brief interview and the standard college placement test.² Which of the four classes an individual student chose to join was not based on any difference in language proficiency but on a personal preference for a particular class time. The make-up of each class can justifiably be described as comparable in respect to language proficiency, age range, and L1 background. Eight subjects were selected for study in each class, although all class members were, in fact, treated in the same way. The students selected were simply those who attended during all 3 weeks of the study. It was only the selected students whose data were analyzed. The classes had begun 3 weeks before data-gathering was started. It was intended that this preresearch period would allow the classes to become established and would lead to more stable populations.

All data were collected during normally scheduled class times. The researcher, who was already familiar to some of the students, took the role of teacher and introduced the tasks as classroom communicative activities. (This was unproblematic, as each of the tasks was closely based on types commonly used in classrooms.) All data were collected using small, unobtrusive dictation machines with no external microphone, which enabled all the dyads to be recorded simultaneously and with the least disruption to normal class routines. In these ways, an authentic classroom setting was preserved as far as possible. It is argued that the use of intact classes under relatively normal conditions minimizes any effects that experimental conditions might have on subjects' performance (Foster, 1994). It makes for greater naturalness in the language produced, not least because of subjects' familiarity with one another (Plough & Gass, 1993). In addition, the inclusion of all class members in the data collection meant that the selected experimental subjects were not alerted to the fact that only their data would be analyzed. The applicability to L2 classrooms of the results of the study is therefore arguably more justified (van Lier, 1988).

Design

Of the four intact classes used in the study, two acted as controls and two as experimental groups. The experimental groups were each divided into two, and in each case half the subjects were randomly assigned the undetailed planning condition and half the detailed planning condition (see Table 2). As all subjects in the experimental groups had 10 minutes of planning time before the task, it was not obvious to either the detailed or the undetailed planners that others in the same

Table 2. Design of the control and experimental groups

| Control Groups | | Experimental Groups | | | |
|----------------|-------------|---------------------|-------------------|---------------------|-------------------|
| Class A | Class B | Class C | | Class D | |
| No planning | No planning | Undetailed planning | Detailed planning | Undetailed planning | Detailed planning |

Table 3. Task order across the groups

| | Control Groups | | Experimental Groups | |
|--------|----------------|-----------|---------------------|-----------|
| | Class A | Class B | Class C | Class D |
| Week 1 | Decision | Narrative | Narrative | Decision |
| Week 2 | Personal | Personal | Personal | Personal |
| Week 3 | Narrative | Decision | Decision | Narrative |

class were responding to a different set of instructions as to how this time might be used.

Each class was visited on three occasions at weekly intervals and on each visit was given one of the three tasks to do. To combat any practice effect, each of the controls and each of the experimental groups did the tasks in a different order (see Table 3). The eventual scores for the Narrative groups, under whichever design condition, were therefore made up of subjects who had done this task in the first week of the study and others who had done it in the third week. The scores for the Decision task were similarly obtained, rendering the two conditions more directly comparable. It was assumed that if there was a practice effect of consistent strength over the three visits of the study, the Personal Information Exchange task, always completed in the second week, would be affected to the same degree as the averaged performance on the other two tasks.

As far as possible, the students worked with the same partner on all of the tasks. At each data-gathering session, all students present in the class, including those not part of the study, were recorded. In this way, subjects were protected from any special pressure that selection for recording might have had upon their performance.

At the end of the 3 weeks, data were transcribed only for 8 subjects in each group who had attended all of the sessions and who had worked with the same partner each time, giving a total of 32 subjects. It was at this stage that the problem with the misclassified student was identified, thus reducing the total sample to 31. Another subject in a control group failed to take her turn in one of the tasks, and an experimental group dyad that had only attended for two of the tasks was nevertheless included to even up the numbers across the four groups. The data set, based on 31 subjects, is otherwise complete.

Analysis

The data were also coded for a range of dependent variables following the rationale provided in the previous section.

Fluency

Reformulations: Either phrases or clauses that are repeated with some modification to syntax, morphology, or word order.

Replacements: Lexical items that are immediately substituted for another.

False starts: Utterances that are abandoned before completion and that may or may not be followed by a reformulation.

Repetitions: Words, phrases, or clauses that are repeated with no modification whatsoever to syntax, morphology, or word order.

Hesitations: Initial phoneme or syllable(s) uttered one or more times before the complete word is spoken.

Pauses: A break of 1.0 second or longer either within a turn or between turns.

Silence total: The sum of pauses in each transcript.

Complexity

Clauses/c-units: Clauses are either a simple independent finite clause or a dependent finite or nonfinite clause. A c-unit is defined as each independent utterance providing referential or pragmatic meaning. Thus, a c-unit may be made up of one simple independent finite clause or else an independent finite clause plus one or more dependent finite or nonfinite clauses.

Syntactic variety: A collection of variables based on verb forms and identifying the tense, modality, voice, and aspect of both finite and nonfinite verbs.

Accuracy

Error-free clauses: A clause in which there is no error in syntax, morphology, or word order. Errors in lexis were counted when the word used was incontrovertibly wrong. In cases of fine decisions of appropriacy, no error was recorded.

Given the focus of the study on planning conditions with three tasks, the general approach to analysis was to perform one-way ANOVAs for each dependent variable for each task, followed by post-hoc tests where *F* values justified this procedure. This approach is most revealing because it is the planning conditions that have the most complex relationship with the various dependent variables. It was also felt important to examine the dependent variables, in the groups outlined, in separate ANOVAs, rather than through a more general multivariate ANOVA. The rationale for the dependent variables presented earlier indicates the distinct role that each contributes. This claim is supported by a factor analytic study of a pooled data set from the present data set and a related study (Skehan & Foster, 1996). This analysis generated a three-factor solution, with the three orthogonal factors clearly identifiable as fluency, complexity, and accuracy, suggesting adequate independence among them.

Table 4. The effects of practice on accuracy

| | No Planning | Undetailed Planning | Detailed Planning |
|---------------|-------------|---------------------|-------------------|
| Narrative | | | |
| Week 1 (0.60) | 0.58 | 0.64 | 0.54 |
| Week 3 (0.64) | 0.64 | 0.69 | 0.61 |
| Decision | | | |
| Week 1 (0.65) | 0.61 | 0.70 | 0.70 |
| Week 3 (0.69) | 0.65 | 0.76 | 0.72 |

Note: Figures represent percentage of error-free clauses.

Table 5. The effects of practice on complexity

| | No Planning | Undetailed Planning | Detailed Planning |
|---------------|-------------|---------------------|-------------------|
| Narrative | | | |
| Week 1 (1.36) | 1.21 | 1.38 | 1.94 |
| Week 3 (1.34) | 1.22 | 1.57 | 1.52 |
| Decision | | | |
| Week 1 (1.33) | 1.20 | 1.41 | 1.55 |
| Week 3 (1.31) | 1.24 | 1.28 | 1.50 |

Note: Figures represent average number of clauses per c-unit.

RESULTS

As indicated earlier (see Table 2), the research employed a design to counterbalance practice effects as the Narrative and Decision tasks were done in both the first and third weeks by different groups of subjects. This design does, however, enable us to probe for practice effects, as one can compare the performance of subjects who completed the Narrative task in week 1 with those who did this task in week 3, and the reverse sequence for the Decision task. The relevant data are shown in Tables 4 and 5. The results focus on the measures for complexity and error, as representative of the wider results available.

None of the complexity figures, for either task, for any of the planning conditions, showed any significant differences. In addition, it is clear from visual inspection of these means that there is no pattern in the data. We can conclude, therefore (and this is interesting in itself), that general task familiarity does not have an influence on the level of complexity achieved. The situation is slightly more complicated for the error measures. No significant differences were found, so we can conclude that the null hypothesis is sustained in this case; that is, familiarity of general task conditions does not lead to improved performance. But two factors suggest that this should be a qualified conclusion. First, for every set of data, there was a higher average score for the third week's performance compared to the first, with this superiority generally being of about 0.05 error-free clauses.³ Second, the sample sizes here were quite small, and it may well be that with larger groups significant

Table 6. *F* values for fluency and task type

| | Personal | Narrative | Decision |
|--------------------------------------|----------|-----------|----------|
| Replacements | 1.59 | 0.37 | 4.48* |
| False starts | 0.15 | 0.13 | 0.84 |
| Reformulations | 1.28 | 0.62 | 0.45 |
| Passives | 0.49 | 2.38 | 1.2 |
| Repetitions/hesitations ^a | 0.32 | 0.79 | 2.99* |
| Pauses | 4.68* | 9.3*** | 12.33*** |
| Total silence | 3.47* | 17.31*** | 21.94*** |

^aThese categories, described separately earlier, were collapsed under one heading for statistical tests.
* $p < 0.5$. *** $p < .001$.

Table 7. Mean scores for fluency and planning condition

| | Pauses | | | Repetitions-Hesitations: Decision |
|---------------------|----------|-----------|----------|--------------------------------------|
| | Personal | Narrative | Decision | |
| No planning | 19.2 | 30.3 | 37.0 | 9.9 |
| Undetailed planning | 10.0 | 15.3 | 17.3 | 20.1 |
| Detailed planning | 11.2 | 8.0 | 17.3 | 17.9 |

| | Total Silence | | | Replacements: Decision |
|---------------------|---------------|-----------|----------|---------------------------|
| | Personal | Narrative | Decision | |
| No planning | 31.8 | 120.3 | 91.4 | 0.56 |
| Undetailed planning | 19.5 | 29.3 | 25.8 | 1.38 |
| Detailed planning | 14.5 | 14.2 | 29.5 | 2.63 |

findings might have been found, in which case one would have to relate familiarity to a slight reduction in the amount of error. In any case, given the counterbalanced nature of the research design, any trend toward a practice effect on errors does not influence the conclusions to be drawn later. As a result, we can now move on to consider the major hypotheses of the study.

Hypothesis 1 proposed that fluency will be greater under planned conditions. Relevant results are presented in Tables 6 and 7. Table 6 gives the separate *F* values for one-way ANOVAs relating the three levels of planning to each of the fluency variables for each of the three tasks. Table 7 gives the actual mean scores for only those comparisons revealed to be significant by the one-way ANOVAs. For example, the *F* value of 3.47 in Table 6 for the measure of total silence for the levels of planning on the Personal task reaches the .05 level of significance, so Table 7 gives the three relevant mean scores of 31.8 seconds (no planning), 19.5 seconds (undetailed planning), and 14.5 seconds (detailed planning). The figures in Table 6 indicate that pauses and total silence achieved significance for all three tasks, with very high levels achieved for the Narrative and Decision tasks. Otherwise, the only

Table 8. Complexity, task type, and planning condition

| | Personal | Narrative | Decision |
|---------------------|----------|-----------|----------|
| <i>F</i> values | | | |
| Clauses/c-units | 8.35 | 9.30 | 15.00 |
| Mean scores | | | |
| No planning | 1.11 | 1.20 | 1.23 |
| Undetailed planning | 1.16 | 1.43 | 1.35 |
| Detailed planning | 1.26 | 1.68 | 1.52 |

***All values are significant at $p < .001$.

significant values were for repetitions–hesitations and replacements for the Decision task.

Table 7 provides the more detailed information for the three planning conditions for those cases where significant *F* values were found. The most striking result here is in terms of the number of pauses and the amount of total silence. The effects are marked and show a consistent effect for planning. In addition, it is striking that, whereas for the Narrative task the different planning conditions themselves (detailed vs. undetailed) each exerted an influence (e.g., pauses: 30.3, 15.3, 8.0), for the Personal and Decision tasks there was only an overall effect for planning versus no planning, the results showing little difference between the two planning conditions (e.g., pauses, Decision: 17.3 vs. 17.3). Perhaps linked to this, the one other result worth commenting on is that for the Decision task; in contrast to the other two, there were significant differences across planning conditions for repetition and hesitation, and for replacement (though not for reformulation) *but in the opposite direction*: That is, planning was associated with *more* repetition, hesitation, and replacement. This combination of findings might suggest that the effects of planning, although very important, are more complex than was expected. Subjects seem to use planning time to capitalize on the use of time-creating devices (Bygate, 1987), as indexed by the effect for repetition and replacement. Such learners seem to use the planning they have done to think “on-line” and possibly to make their discourse more naturalistic. We can conclude, therefore, that Hypothesis 1 is supported for pauses and total silence. Variables such as repetition and replacement clearly merit further research.

Hypothesis 2 stated that planning would be associated with greater complexity, with complexity indexed by the ratio of clauses divided by c-units. Following the sequence used with the fluency measures, Table 8 shows first the *F* values from one-way ANOVAs for the subordination measures (e.g., 8.35 being the *F* value for the one-way analysis of the three planning conditions for the Personal task). Table 8 shows next the mean scores for the three planning conditions on this task—for example, 1.11, 1.16, and 1.26 as the three mean scores dependent on the *F* value of 8.35.

The results in Table 8 suggest a very strong effect for the planning condition: Detailed planning produced significantly more subordination than undetailed planning, which in turn produced significantly more subordination than no planning.

Table 9. Syntactic variety measures

| | Personal | Narrative | Decision |
|---------------------------------------|----------|-----------|----------|
| <i>F</i> values | | | |
| Nonsimple present tenses | 0.06 | 0.10 | 0.80 |
| Nonsimple past tenses | 3.46* | 5.12** | 1.93 |
| Modal total | 0.56 | 0.99 | 0.49 |
| Conditional total | 1.42 | 1.45 | 2.75 |
| Mean scores for nonsimple past tenses | | | |
| No planning | 0.0 | 0.0 | |
| Undetailed planning | 0.0 | 1.13 | |
| Detailed planning | 0.25 | 1.17 | |

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

Range tests using Duncan's procedure (Norusis, 1990) indicate that within each task all pairwise comparisons within each task are significantly different from one another with the exception of the no planning (1.11) and undetailed planning (1.16) conditions for the Personal task, which were not. Hypothesis 2 is accordingly confirmed.

Hypothesis 3 stated that planning would be associated with greater structural variety. The relevant results are shown in Table 9. *F* values are shown for the one-way ANOVAs that examined the effects of planning conditions for each variable for each task; for example, 5.12 is the *F* value for the three planning conditions for the Narrative task and the dependent variable of nonsimple past tenses. The actual mean scores for the three planning conditions are shown for the two cases (nonsimple past tenses on the Personal and Narrative tasks) in which significant *F* values were found. These mean scores are necessarily low because opportunity for use of the different structures is unavoidably limited in scope and because they are given in the form of proportion scores.

The significant results all involve use of the past tense. The measures examine the number of uses of past-tense forms other than the simple past, that is, those that could be argued to be more interesting and challenging forms, and whose use reflects learners who are more willing and able to take risks, restructure, and use forms at the borders of interlanguage competence. The results show that for the Personal and Narrative forms significance is achieved, whereas for the Decision task, in which the *F* value of 1.93 produces a probability value of only $p < .16$, significance was approached but not attained. The detailed mean scores then show that planning is associated with greater variety of past-tense usage. No significant differences were found for greater variety of present-tense usage (i.e., forms other than the simple present), modals, or conditionals. Hypothesis 3 therefore receives some support, but only at a fairly weak level.

Hypothesis 4 stated that planning would be associated with greater accuracy. Relevant results are presented in Table 10. Only one measure is presented here, the percentage of error-free clauses. The three *F* values are shown first, followed by detailed means scores.

Table 10. Error as a function of task type and planning condition

| <i>F</i> -values | Personal | Narrative | Decision | Mean |
|---|----------|-----------|----------|------|
| <i>F</i> values from one-way ANOVAs | 2.46* | .69ns | 5.73** | |
| Mean error scores by planning condition by task | | | | |
| No planning | 0.64 | 0.61 | 0.63 | 0.63 |
| Undetailed planning | 0.76 | 0.65 | 0.73 | 0.71 |
| Detailed planning | 0.69 | 0.58 | 0.71 | 0.66 |
| Mean score | 0.70 | 0.61 | 0.69 | |

* $p < .05$. ** $p < .01$. ns = not significant.

The central finding here is that there are significant differences in the proportion of error-free clauses that were used under the different planning conditions, but these differences show a complex pattern. The *F* ratio of 5.73 for the Decision task is significant at the .01 level, with Duncan range tests demonstrating that the no planning condition was significantly different from the two planning conditions but that the two types of planning did not produce significantly different effects. The *F* value for the Personal task just attains the .05 level of significance. Detailed tests indicate that the no planning condition and the undetailed planners are significantly different from one another. No other comparisons are significant. Significance is not attained for the Narrative task. These results provide limited support for Hypothesis 4.

Hypothesis 5 proposed that the effects of planning would be stronger as planning is carried out at a more detailed level, that is, that there would be a basically monotonic relationship over the three planning levels, with the detailed planning condition producing stronger effects than the undetailed condition, which in turn would have stronger effects than the no planning condition. This hypothesis can be explored by drawing upon the data already presented in Tables 6–10. Regarding the fluency measures of pauses and silence, the no planning group always produced the least fluency. Detailed planning only led to greater fluency than undetailed planning for the Narrative task. For the Personal and Decision tasks variation in planning produced little difference. The findings for complexity, in contrast, essentially supported Hypothesis 5 across the tasks, with the detailed planners always producing significantly greater complexity than the undetailed planners, who in turn outperformed the no planning group (with only the comparison for the no planners vs. the undetailed planners failing to reach significance). The syntactic variety results, which provided only limited support for the more specific Hypothesis 3, patterned similarly to the fluency measures, with little difference between the two planning conditions.

Finally, we need to consider the accuracy results. For each task it is noteworthy that the undetailed planners produced the most accurate level of performance. This is strikingly so in the case of the Personal task (0.76 vs. 0.64 and 0.69) and the Narrative task (0.65 vs. 0.61 and 0.58) but only marginally true in the case of the

Decision task (0.73 vs. 0.71 and 0.63). In addition, the detailed planners produced a level of accuracy that was greater than the no planning condition for the Personal and Decision tasks, but the no planning condition produced more accurate performance than detailed planning in the Narrative task. These comparisons do not attain significance in specific pairwise comparisons, but the trend in the results is clear.

Two aspects of these findings merit additional comment. First, these results contrast with those of Crookes (1989) in that here planning *is* associated with a greater degree of accuracy, at least some of the time. It does appear, in other words, that some of the value of having planning time available is channeled into achieving more error-free language. Second, the results also contrast somewhat with those for subordination, in which there was a monotonic relationship with planning condition and in which the detailed planners produced the most subordination on all three tasks. With accuracy, the trend is toward the undetailed planners being the most accurate. We will return to this finding; at this point, we have to conclude that Hypothesis 5 is only supported to a limited degree and that the results obtained reveal a more complex picture than was predicted.

Finally, Hypothesis 6 proposed that the effects predicted in Hypotheses 1–5 would be greater for the more cognitively demanding tasks, that is, those based on less familiar information and more cognitive processing. In the present case, this contrasts the Personal task with the Narrative and Decision tasks. Once again, the hypothesis can be explored by going over the data presented in Tables 6–10. There is some support for this hypothesis as shown by the fluency measures for pauses and total silence in Table 7. The fluency measures for the Personal task reduced by less than do those for the Decision task and, most of all, the Narrative, where total silence decreased from 120.3 seconds (no planning) to 14.2 seconds (detailed planning), an enormous change compared to a reduction of around 17 seconds for the Personal task (31.8–14.5).

A similar picture emerges for the complexity measures. The Personal task always produced the least subordination, perhaps reflecting the linguistic demands that were made to express directional meanings, as well as the succinctness that may result from well-organized knowledge. Furthermore, the difference between the no planning and detailed planning conditions for the Personal task is only 0.15 (i.e., 1.26–1.11). The corresponding figures for the Decision and Narrative tasks are 0.29 and 0.48, respectively, appreciably larger figures. Planning, in other words, does not operate in the same way with all tasks. It would be valuable to probe which task qualities—for example, use of less familiar knowledge or greater on-line processing—account for these differential effects.

The measures of syntactic variety were not strong, in any sense, and provide no clear support for Hypothesis 6. The use of nonsimple past tenses to refer to past time generated the only significant results. Inspection of the actual mean scores provides no supportive evidence here, and indeed the fact that the Personal task generated one significant *F* value when the Decision task did not seems to argue against Hypothesis 6. Similarly, the results for accuracy do not support Hypothesis 6. The greatest difference between a no planning and a planning score is that between no planning and undetailed planning Personal scores. The gain scores for accuracy on

the other two tasks are lower and, indeed, the detailed planners on the Narrative produce lower accuracy scores than the no planners.

We can conclude, therefore, that Hypothesis 6, which can be broken down into subhypotheses for the different dependent variables, does not yield a unified picture. It is supported by the complexity results and partially supported by the fluency results. In contrast, the syntactic variety results do not support the hypothesis, whereas the results for accuracy, if anything, run counter to it. This mixed picture will be discussed in the next section in terms of tradeoff effects that result from processing capacity limitations.

DISCUSSION

The general perspective in this article is that attention is limited in capacity and that its use to achieve one goal can reduce the capacity remaining for the achievement of other goals. Thus, the framework proposed in Table 1 was intended to enable the organization of task-based instruction to minimize its dangers (excessive focus on meaning) while maximizing the chances that its advantages will be realized (accurate and complex form). The framework also generates hypotheses that are testable, and this article has focused on one phase (pretask activities) and one area within that phase (planning), as well as on the nature of the tasks themselves.⁴

Regarding task difficulty, it was proposed earlier that the Personal task would be the easiest, the Decision task the most difficult, with the Narrative closer to the Decision than to the Personal. The results obtained only partly fit into this analysis. The Personal task indeed produced much more fluent performance, as indexed by the average number of pauses and the average amount of silence, with the differential between the tasks being interestingly narrowed for the more planned conditions. This applies dramatically to the total silence figure for the unplanned Narrative task and suggests that being able to draw upon familiar, ready-encoded information does promote a greater degree of fluency and a corresponding lack of need for planning.

The pattern changes for the figures on average error. Here, the Personal and Decision tasks lead (average error proportions of 0.70 and 0.69, respectively), whereas the Narrative task produces the lowest overall accuracy level (0.61). But these rankings change when we consider the average complexity figures. The least complex language is used with the Personal task (1.18 clauses per c-unit), the Decision task produces more complex language (1.32), and the Narrative task produces the most complex language of all (1.44). But these average figures disguise some important interactions between task type and planning condition. First of all, it is noteworthy that the no planning condition produces a relatively even level of performance across the tasks. The range in scores across the tasks is only 0.03 for accuracy and 0.08 for complexity. In contrast, the two planning conditions lead to much more uneven figures. With accuracy, the Personal task and the Decision task produce clear (and fairly similar) gains, whereas the degree of improvement for the Narrative task is less marked. A contrasting picture emerges with the complexity measures. The smallest improvement shown as a function of planning is for the Personal task (i.e., the opportunity to use more time does not seem here to confer

much advantage for complexity). The Decision task is influenced appreciably by planning, and the Narrative task is influenced most of all.

Generalizing from these results, what seems to be at issue here is that each of the tasks produces a pattern of results that basically reflects task properties. The Personal task generates a greater degree of accuracy without achieving much complexity. The Narrative task produces the highest level of complexity, suggesting a greater need to use precise and extending language, but this gain seems to be at the expense of accuracy, as this highest complexity measure (1.68) is associated with the lowest accuracy score (0.58). Finally, the Decision task seems to occupy an intermediate position, producing useful levels of accuracy *and* complexity. The accuracy level, it will be recalled, is comparable to that obtained with the Personal task (and above that obtained for the Narrative task). Complexity, in contrast, although never as great as the Narrative task, is higher than the Personal task and shows a more marked effect for planning. This suggests that the combination of task type and planning condition provides scope for complexity to operate as a viable goal without compromising accuracy.

More broadly, what seems to be happening here is that subjects were operating under some information-processing pressure: They had to allocate attention to particular goals at the expense of other goals. The pattern of results, in other words, supports *tradeoff effects* between complexity and accuracy, a point developed more fully in Skehan and Foster (1996). Decision task subjects under planning conditions were most able to combine goals simultaneously. Other tasks seemed to lead to a greater degree of sacrifice of one goal or another. The implication, in other words, is that task selection and task implementation can lead to systematic and selective influences upon the nature of the ensuing proficiency.

We turn next to examining the effect of the planning variable in its own terms. We have seen briefly that planning has an effect upon fluency. This effect is of a monotonic nature for the Narrative task, but there is only a binary distinction for the other two tasks (i.e., the detailed and undetailed conditions do not lead to different patterns of pausing and total silence). Clearly, subjects required to wrestle with the complexity of the story to be told in the Narrative task receive additional benefit from the structure that the detailed planning condition provides. Possibly, with tasks that are most demanding cognitively, the planning suggestions provided to subjects enable them to make more inroads into a problem they might otherwise not be able to analyze so effectively. Interestingly, the one other fluency effect for planning is with the Decision task and the variables of repetition/hesitation and replacement. Planning produces more of these, rendering the discourse, in one sense, less fluent. But we interpret this finding to suggest naturalness of discourse, in that native speakers also engage in a greater use of such linguistic behaviors to enable them to cope with the problem of real-time. In other words, it may be that the subjects under the unplanned condition, who would benefit most from the use of techniques to protect time for ongoing planning, are unable to exploit them. In contrast, the planned condition subjects, already favored by the experimental condition they received, are able to capitalize upon their clear ideas and language and to make

things even easier by exploiting repetitions and replacements to create a natural buffer against time pressure.

The complexity and error measures make for an interesting contrast. The complexity values provide a very clear picture: The relationship is linear, and it does appear that detailed planners have responded to the suggestions given to them by producing more subordinated language. At present, we cannot say whether this is the result of incorporating more complex forms into speech consciously, or the detailed planning has reduced cognitive load to enable greater challenges to be attempted during ongoing discourse, or the effects of the (detailed) planning have been cognitive, and so pushed learners to find language to express more complex ideas. We suspect the last of these, a proposal consistent with the reactions of the subjects, who thought the Narrative task most difficult but produced the most complex language while doing it. Further research will be necessary to try to establish what the operative causal factors are.

But what is even more interesting still is that this pattern of results is not repeated for the accuracy measures. Here, the accuracy level order for the Personal and Decision tasks is no planning < detailed planning < undetailed planning, whereas for the Narrative task the order is detailed planning < no planning < undetailed planning. In all cases, though, the most accurate performance is for the undetailed planners.⁵ This pattern of results was not predicted. First of all, we need to recall that the evidence on the influence of planning on accuracy is mixed. Although Ellis (1987) reported such an effect, it was for a study that focused on only one language form (the past tense, so that no complexity dimension was involved) and for which the experimental conditions (talking, alone, into a tape recorder) might be thought to favor accuracy. The most comparable study (Crookes, 1989) reported a complexity effect, but no significant differences for accuracy. A generally similar pattern of results is reported in Alvarez-Ossario (1994). So it is already striking that there is an accuracy effect in the present study, even though statistically it is not as strong as that for complexity or fluency. But even more surprising is that the detailed planners, who were expected to exploit the guiding instructions to facilitate generalized improvement, did not seem to do so.

In trying to account for the discrepancy with Crookes (1989) and Alvarez-Ossario (1996), two potential explanations present themselves. First of all, the contrast in findings may be due to the nature of the measures used. Crookes (1989) looked for specific effects on a wide range of variables, whereas the present study used a more global measure based on all contributory influences on error and correctness. Crookes (1989) seemed to be testing for specific increases in accuracy, whereas we were more interested in detecting a more generalized accuracy effect. It may simply be, therefore, that the use of a dependent variable such as error-free clauses is more sensitive to an overall accuracy effect, a proposal perhaps consistent with the weaker finding in this area than for complexity. A second explanation, however, focuses on the tasks themselves. Crookes' (1989) tasks, for example, a LEGO assembly, were classroom tasks and did not have much connection with the subjects' real lives. They were, quite simply, devices for the transmission of neutral and arbitrary

information. The tasks in this study contrast with these to some degree. This is clearly the case with the Personal and Decision tasks. In the first case, students described an aspect of their actual lives, whereas in the second they had to draw upon their own system of values and beliefs. There is evidence (Tarone, 1985) indicating that discourse salience can influence the level of accuracy on a task, particularly when syntax and morphology have discursal value to signal meaning. Although this is, of course, speculative, it may be the case that subjects are drawn toward a greater degree of precision and accuracy when they can use planning conditions to invest the language they use in tasks with more personal significance. In passing, it should be noted that the Narrative task is less effective at meeting this criterion of real-world connection, as there is obviously an arbitrary, detached quality to producing an imaginary story based on a set of (relatively) unrelated photographs. It is relevant, in this respect, that the changes in accuracy were least for this task, with even a reduction of accuracy level for the detailed versus the undetailed planners. In any case, the two explanations for the discrepancies in research findings are post hoc and will have to await further research directed specifically at this issue before they can be resolved.

Even so, two points remain that do require discussion: the nature of the relationship between the accuracy and complexity results, and the precise role of detailed and undetailed planning. In reality, these points can be accounted for together. What seems to be happening is that the goals of complexity and accuracy compete for limited information processing resources and that what is achievable depends on the precise allocational decisions made. Certainly the joint levels of accuracy and complexity achieved by the undetailed planners reflect progress being made toward both goals simultaneously (as well as toward fluency). In that respect, the psycholinguistic resources available seemed to cope adequately with the pressures that had been manipulated. But the next condition, detailed planning, seemed to lead to a level of pressure too great for subjects to cope with all aspects of performance, with the result that prioritizing decisions had to be made. Consistent with Crookes (1989), complexity became the preeminent goal, but accuracy suffered to some degree, reducing to a point midway between the undetailed and nonplanning conditions for the Personal task, below the no planning conditions for the Narrative task, and only slightly inferior to the undetailed planning for the Decision task.

Turning finally to the explanation for the excessive pressure on processing resources that is produced by the detailed planning condition, the most plausible account would seem to be that detailed planning causes subjects to structure their subsequent task contributions in a more ambitious manner. This may have been in terms of propositional content and discursal development, such that the planning time was channeled into the production of more complex ideas, which then pushed learners to levels of syntax beyond their level of comfortable control or, alternatively, whose execution was so demanding on available resources that, through the prioritization of more interesting ideas, accuracy was squeezed out. Alternatively, the detailed planning time was used to access more "cutting edge" language that, precisely because of its developing nature, was less controlled. Further research

will have to probe which of these interesting alternatives was the more powerful influence.

CONCLUSION

Two broad implications can be drawn from the research reported here. First of all, there is the issue of the framework within which research can be most usefully conducted in the area of task-based instruction. This study has drawn on the framework from Table 1, which suggests a number of factors, organized as pre-, mid-, and posttask stages, that may influence the balance struck in task-based instruction among goals of fluency, complexity, and accuracy. It has shown that pretask activities can have a significant impact on the ways that tasks are done. But what the framework also does is suggest other areas that can be researched to produce a better overall understanding of how tasks can be most effectively used pedagogically. Simply building on the present results, it is clear that, interesting though they are, and immediately useful as they may be for pedagogic decisions, they raise as many questions as they provide answers. Although it seems likely that tradeoff attentional effects are central, the exact patterning of results was not completely predicted, and it is clear that additional research is now urgently necessary into the precise effects of planning time on the competing goals of complexity and accuracy, especially. Planning time has a major impact, but it would seem that if we could learn more about this area it would be possible to predict more effectively the exact balance between these different goals. In this way, it would be possible to set longer term pedagogic goals in which cycles of synthesis and analysis are engaged and areas of interlanguage are developed and then integrated into fluent performance. To achieve this, clearer operationalizations of planning time, as a construct, are necessary, as well as the use of research methodologies, such as immediate retrospection, that provide more detailed information on which planning operations were undertaken and what the differential effects of such operations might be.

The framework from Table 1, however, encompasses more than simply pretask activities. It also suggests that midtask and posttask activities are important. Skehan and Foster (1996) researched posttask activities, and this research suggests that the knowledge or threat that a posttask repeat public performance of the task being done is likely to be required increases the focus on accuracy. The framework also suggests that midtask manipulations may also influence the balance among the goals of fluency, complexity, and accuracy. Research into the injection of surprise elements, or the use of different degrees of visual support, may well be productive (Foster & Skehan, 1996).

Finally, the three tasks used in the present study—a Personal Information Exchange task, a Narrative task, and a Decision-Making task—yielded interestingly contrasting results. These three tasks were chosen because they vary the importance of background knowledge and general familiarity, on the one hand, and inherent structure, on the other. The results obtained were consistent with the hypothesized task properties. But these conclusions are based on just three tasks. It is also impor-

tant to conduct research on other examples of these three task types, as well as on other task types, and to explore the predictability of the language characteristics associated with such tasks. This is essential if one is to be able to use tasks whose worth can be predicted in advance rather than only observable after task completion. In this way, perhaps, a library of tasks, of known qualities, can be drawn upon as a resource by the teacher interested in task-based instruction. These issues are explored more fully in Skehan and Foster (1996).

(Received 25 August 1995)

NOTES

1. Before the study proper began, a number of candidate tasks were piloted at a neighboring college in order to determine for each of the three types which task was the most productive of language and the most enjoyable to do. In this way, the risk of basing the research project on unpopular and unproductive tasks was reduced.

2. Subsequently, one subject had to be discarded when it was realized that the college had not applied its own selection policy to him. Despite a low test score, he had been placed in one of the four groups simply for scheduling convenience reasons. When the lower level of his performance was noticed, his results were removed from the data set.

3. In any case here, even if there were an effect that has not been detected in the present study, the underlying cause may be task familiarity, or simply improvements in proficiency over time. The students were following courses of instruction, and it may be that this instruction was having an effect. Even so, it must be borne in mind that the effect, in such a case, was on accuracy, not on complexity.

4. See Skehan and Foster (1996) and Foster and Skehan (1996) for studies on posttask and during-task influences, respectively.

5. This pattern of results has been replicated in a subsequent study (Foster & Skehan, 1996).

REFERENCES

- Alvarez-Ossario, L. (1996). *The effect of planning in L2 Spanish narratives*. Unpublished manuscript, University of Hawaii at Manoa.
- Anderson, J. R. (1995). *Learning and memory*. New York: John Wiley.
- Aston, G. (1986). Trouble-shooting in interaction with learners: The more the merrier? *Applied Linguistics*, 7, 128–143.
- Brock, C. (1986). The effects of referential questions on ESL classroom discourse. *TESOL Quarterly*, 20, 47–59.
- Brown, G., Anderson, A., Shilcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.
- Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.
- Candlin, C. (1987). Towards task-based language learning. In C. Candlin & D. Murphy (Eds.), *Language learning tasks* (pp. 5–22). Englewood Cliffs, NJ: Prentice-Hall.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367–383.
- Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study on SL relativization. *Studies in Second Language Acquisition*, 13, 431–469.
- Duff, P. (1986). Another look at interlanguage talk: Taking task to task. In R. Day (Ed.), *Talking to learn* (pp. 147–181). Rowley, MA: Newbury House.
- Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, 9, 12–20.
- Foster, P. (1994). *Discoursal outcomes of small-group work in an EFL classroom*. (Working Papers in English Language Teaching No. 2). London: Thames Valley University.
- Foster, P., & Skehan, P. (1996). *The effects on accuracy and complexity of planning and mid-task manipulation*. Manuscript in preparation.
- Harmer, J. (1991). *The practice of English language teaching*, 2nd ed. London: Longman.

- Harrison, A. (1986). Assessing text in action. In M. Portal (Ed.), *Innovations in language testing* (pp. 72–96). Windsor, UK: National Foundation for Educational Research.
- Hunt, K. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732–739.
- Long, M. (1989). Task, group, and task–group interaction. *University of Hawaii Working Papers in English as a Second Language*, 8(20), 1–26.
- Long, M., & Crookes, G. (1991). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26, 27–55.
- Long, M., & Crookes, G. (1993). Units of analysis in syllabus design: The case for task. In G. Crookes & S. Gass (Eds.), *Tasks in a pedagogical context: Integrating theory and practice* (pp. 9–54). Clevedon, UK: Multilingual Matters.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Clevedon, UK: Multilingual Matters.
- Norusis, M. (1990). *SPSS PC user's manual, Version 4*. Chicago: Statistical Package for the Social Sciences.
- Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communicative tasks for second language instruction. In S. Gass & G. Crookes (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9–34). Clevedon, UK: Multilingual Matters.
- Plough, I., & Gass, S. (1993). Interlocutor and task familiarity. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 35–56). Clevedon, UK: Multilingual Matters.
- Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.
- Rivers, W. (1981). *Teaching foreign language skills*. Chicago: University of Chicago Press.
- Robinson, P. (1995). Attention, memory, and the noticing hypothesis. *Language Learning*, 45, 283–331.
- Robinson, P., Ting, S. C.-C., & Urwin, J. J. (1995). Investigating second language task complexity. *RELC Journal*, 26, 62–79.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 17–46.
- Schmidt, R. (1994). Deconstructing consciousness: In search of useful definitions for applied linguistics. *AILA Review*, 11, 11–26.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Skehan, P. (1992). *Strategies in second language acquisition*. (Working Papers in English Language Teaching No. 1). London: Thames Valley University.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38–62.
- Skehan, P., & Foster, P. (1996). *The influence of post-task activities and planning on task-based performance*. (Working Papers in English Language Teaching No. 3). London: Thames Valley University.
- Tarone, E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, 35, 373–403.
- van Lier, L. (1988). *The classroom and the language learner*. London: Longman.
- VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287–301.
- Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London: Collins.
- Willis, J., & Willis, D. (1988). *Collins COBUILD English Course: Book 1*. London: Collins.