SSLA, **20**, 83–108. Printed in the United States of America.

THE EFFECTS OF DIFFERENT LENGTHS OF TIME FOR PLANNING ON SECOND LANGUAGE PERFORMANCE

Uta Mehnert

Thames Valley University

This article reports on a study that investigated the effect of different amounts of planning time on the speech performance of L2 speakers. Subjects were 4 groups of learners of German (31 in total) performing 2 tasks each. The tasks varied in the degree of structure they contained and the familiarity of information they tapped. The control group had no planning time available; the 3 experimental groups had 1, 5, and 10 minutes of planning time, respectively, before they started speaking. Results show fluency and lexical density of speech increase as a function of planning time. Accuracy of speech improved with only 1 minute planning but did not increase with more planning time. Complexity of speech was significantly higher for the 10-minute planning condition only. No significant differences were found for the effect of planning on the different tasks. This study employed various general and specific constructs for measuring fluency, complexity, and accuracy of speech. The interrelationships and qualities of these measures are also investigated and discussed.

Speech production is seen by most researchers as a planned process. Clark and Clark (1977) divided speaking into two types of activity: planning and execution. Similarly, Færch and Kasper (1983) distinguished between the planning phase and the execution phase. They also made a distinction between the planning process and the plan itself, which will control the execution phase. In the planning phase, the language user scans the linguistic repertoire and selects rules and items by means of which a communicative goal can be reached and combines them into plans at the different hierarchical levels (Færch & Kasper, 1986).

The author wishes to thank Professor Peter Skehan as well as four anonymous *SSLA* reviewers for their helpful comments on an earlier draft of this article.

© 1998 Cambridge University Press 0272-2631/98 $9.50

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

Færch and Kasper (1983) further argued that planning can be more or less conscious and that in L1 production the planning process is normally subconscious and highly automatic. This means that for many recurrent, standardized situations, speakers have available to them ready-made plans that contribute to reducing the processing load. These ready-made plans can also be modified or integrated into larger newly constructed plans to fit the specific requirements of the communicative goal. Creatively constructed plans can themselves be stored in the speaker's declarative knowledge as ready-made plans.

Many L2 speakers, especially at the beginner level, have to creatively construct plans for most situations because ready-made plans are not yet available, which means that the activation of procedures (e.g., lexical retrieval and morphological marking) requires a high degree of cognitive control. In order to reduce cognitive strain or processing load, it has been suggested (Crookes, 1989; Foster & Skehan, 1996; Skehan, 1996) that advance speech planning is an effective way for L2 learners to handle communicative strains and pressures, thereby improving their performance.

In the last few years, a number of studies have reported on the effects that advance speech planning has on the performance of L2 learners. These studies vary in a number of respects, including the motivation for studying speech planning and the aspects of L2 learners' speech performance that were analyzed. However, most of them operationalized advance speech planning as time given in advance of task performance and compared L2 learners' speech performance under two conditions: an unplanned condition, in which subjects had no time to plan their speech, and a planned condition, in which they had time to plan their speech.

Crookes (1989) investigated planning as a possible cause for interlanguage variation and development and reported on an experiment in which Japanese learners of English as a second language performed two monologic production tasks with and without time for planning. Under the planned condition, subjects had 10 minutes available and were instructed to plan their speech in terms of words, phrases, and ideas. Subjects were asked to make written notes in order to ensure that they did in fact engage in planning. Crookes (1989) found that under the planned condition interlanguage productions were more complex, measured in terms of words, number of subordinate clauses, and sentence nodes (s-nodes) per utterance. With regard to accuracy, measured in terms of length of error-free T-units and number of error-free T-units per utterance, differences were not statistically significant.

Working within the context of task-based instruction, Foster and Skehan (1996) showed a consistent effect for planning on fluency (number of pauses and total silence), complexity (subordination and variety of past tense usage), and accuracy (error-free clauses) of L2 learners' speech performance. Using three different tasks, they also found evidence of an interaction between planning and task type; the effects of planning, in particular on fluency and complexity, were greater for the more cognitively demanding tasks. Foster and Skehan concluded that planning is a useful pretask activity that can influence the way tasks are done and thus affect L2 learners' level of performance.

Wigglesworth (in press) was concerned with the provision of planning time in a

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

testing situation. She found that subjects benefited from planning periods as brief as 1 minute; that is, their performance improved in terms of complexity (subordination), fluency (self-repairs), and accuracy (verb morphology). She also found selective effects of planning dependent on task type as well as proficiency level. High-proficiency subjects benefited from planning more than did low-proficiency subjects, and the benefit was greatest on the more difficult tasks.

Because the three studies discussed above employ different tasks and different variables to measure speech performance, it is impossible to make any precise comparisons as to the effect different amounts of planning time have on L2 learners' performance. In particular, it would be useful to examine whether increasing amounts of planning time have a systematically increasing effect on L2 learners' speech performance (cf. Wigglesworth's [in press] use of 1 minute planning, whereas Crookes [1989] and Foster and Skehan [1996] use a 10-minute planning period) and whether any such effect can be observed across different aspects of speech performance. The study reported here will attempt a systematic investigation of the influence of amount of planning time on learners' speech performance and thus provide potentially valuable findings that subsequent studies can build upon.

Another issue that needs to be addressed is the question of how to measure L2 learners' speech performance. Skehan (1996) put forward fluency, accuracy, and complexity as three goals in foreign-language learning. Following this, Foster and Skehan (1996) and Wigglesworth (in press) analyzed subjects' speech performance in terms of fluency, accuracy, and complexity, but used different variables to measure performance in these three areas. We therefore need to consider in more detail the various measures and their usefulness in order to be able to compare studies more effectively.

With regard to fluency, pretask planning is thought to reduce the amount of online speech planning needed, which means that more planned speech will be more fluent than less planned speech. Increased fluency is reflected in fewer or shorter pauses, or both. Foster and Skehan (1996) included in their analysis pauses of 1 second and over, and they found that under planned conditions, subjects paused significantly less frequently. The amount of total pausing time was also significantly smaller for planners as compared to nonplanners. A measure closely linked to number of pauses is mean length of run. It indicates the average number of syllables between pauses. Another good overall indicator of fluency that also takes into account the time used for pausing is speech rate (Möhle, 1984; Raupach, 1980; Towell, 1987), which indicates the number of syllables produced on average per minute. In order to capture the influence of varying numbers of repetitions, reformulations, and replacements, a second measure of speech rate will be determined by excluding from the count all phrases, words, and syllables that were subsequently repeated, reformulated, or replaced.

It is assumed that under planned conditions subjects attempt to express more complex propositions, resulting in greater subordination. At the same time, planning provides subjects with the opportunity to fill out propositional slots (Ellis, 1984), resulting in longer T-units. Some researchers (Crookes, 1989; Wigglesworth, in press) used subordination (i.e., the number of subordinate clauses) to measure language

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

complexity. Analyzing L2 learners' speech production, Crookes (1989) also used s-nodes and concluded that this provides a broader measure of complexity than subordinate clauses. In his study, s-nodes covered all untensed verb forms. In Brock's (1986) study, infinitives and gerunds, as well as tensed verbs, but not modals, were taken to signal an underlying s-node. Following a similar line, Foster and Skehan (1996) measured complexity as the number of clauses, including both independent and dependent clauses, the latter being either finite or nonfinite.

In many studies, accuracy has been measured by *specific* measures, such as past tense morphemes (Ellis, 1987) and plural *-s* (Crookes, 1989; Wigglesworth, in press). In many cases, these measures did not reveal significant differences between planning conditions. However, just because the specific measures chosen in these studies did not reveal any differences does not necessarily mean that there were no differences in accuracy with regard to other language features. It would thus probably be more effective to use more *general* measures that capture the full extent of any difference by determining the overall language accuracy. Foster and Skehan (1996) proposed the proportion of error-free clauses as a more appropriate indicator of speech accuracy. They found significant differences in the use of error-free clauses between planners and nonplanners for some of their tasks. However, by using error-free clauses, no distinction is made between clauses with one error and clauses with multiple errors. Error-free clauses may, therefore, not be a particularly good measure in some cases, such as with learners of relatively low proficiency or learners of a language like German, which is highly inflected for nouns, articles, pronouns, adjectives, and verbs. Even with advanced students of German, errors due to wrong inflection amount to almost 30% of all errors (Rogers, 1984). Other measures of overall accuracy that take account of this are number of errors per clause (Bardovi-Harlig & Bofman, 1989) or, since clauses can be of different lengths, number of errors per 100 words.

A number of studies show that there are differences in lexical density between spoken and written discourse (Halliday, 1985; Ure, 1971), as well as between different types of spoken discourse (O'Loughlin, 1995; Zora & Johns-Lewis, 1989). Ure (1971) and Zora and Johns-Lewis (1989), as well as results obtained by O'Loughlin (1995), suggested that the degree to which discourse is planned may be a possible cause of differences in lexical density between different discourse forms. In order to test this hypothesis, a measure of lexical density was included in the present study.

Lexical density measures the relationship between lexical and grammatical items in a text by calculating the ratio of the number of lexical items to the total number of items. Halliday (1985) argued that the use of weighted lexical density gives a truer, more fine-grained picture of lexical density. Weighted lexical density distinguishes between low- and high-frequency lexical items, the latter being items that either occur commonly in the language in general or more than once in an individual text because, as he argued, repetition reduces the effect of density.

The study reported here was designed to test the influence of amount of planning time on speech performance, addressing the following questions: (a) Does more planning time result in improved performance; that is, does longer planning time

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

produce additional benefits? If yes, (b) does performance increase in proportion to the amount of planning time made available? and (c) will any increase in performance be the same for fluency, complexity, and accuracy? In addition, this study will also look at the influence of planning on different tasks and the usefulness of various general and specific measures of fluency, complexity, and accuracy, as well as the place of lexical density within this framework. The following hypotheses can be formulated:

Hypothesis 1. Speech produced under planned conditions will be more fluent than that produced under unplanned conditions. This follows from Foster and Skehan (1996), who found that planners pause significantly less frequently and that the amount of total silence in their performance is significantly smaller than that of nonplanners.

Hypothesis 2. Under planned conditions, learners will produce more complex language. Foster and Skehan (1996) reported that planned speech was associated with a greater number of clauses per communication unit (c-unit). Crookes (1989) found a greater number of subordinate clauses in planned speech; the difference was, however, not statistically significant. He also found longer utterances and longer subordinate clauses.

Hypothesis 3. Learners' speech under planned conditions will be more accurate. This is based on results reported by Ellis (1987), Foster and Skehan (1996), and Wigglesworth (in press). Ellis found greater accuracy of regular past tense forms under the planned conditions. Foster and Skehan (1996) reported a greater proportion of error-free clauses under planned conditions for some of their tasks. Wigglesworth (in press) found verb morphology to be more accurate under the planned condition. In contrast, Crookes found neither greater use of error-free T-units under the planned condition nor significant differences for plural *-s* accuracy.

Hypothesis 4. Speech produced under planned conditions will show signs of greater density of language. This hypothesis is based on research reported by Ure (1971) and O'Loughlin (1995) that suggests that more planned speech will have a higher lexical density.

Hypothesis 5. The effects of planning described in Hypotheses 1–4 will increase in line with the amount of planning time made available.

Hypothesis 6. The effects of planning will be greater for the cognitively more difficult task. This follows from Foster and Skehan (1996), who found that more difficult tasks yield more complex language. Wigglesworth (in press) also found greater effects of planning for more difficult tasks with respect to subordination and accuracy of verb morphology.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

METHOD

Tasks

Because all tasks, for practical reasons, were to be performed in individual recording booths of a language laboratory, it was considered important to design tasks for which the use of the language laboratory would not constitute too artificial an environment. Two tasks were designed that required subjects to leave a message on a friend's answering machine. It was felt that this setting would fit well with the language-laboratory environment with the subjects speaking to a machine and having no direct and immediate contact with anyone, as well as resembling a situation in which advance speech planning is conceivable and possibly frequently adopted in real life, at least by L2 learners. Both tasks required the oral production of monologues.

In the instruction task, the subject was required to leave a message on an answering machine telling his German friend that he could not meet him at the airport, explaining how he could get from the airport to the university, and when and where he would meet him there. In the exposition task, the subject was required to apologize to two German fellow students and come up with a good excuse for why he did not meet them the day before as arranged. The story had to contain six specified words. In this way, the two tasks bear some similarity to the personal and narrative tasks used by Foster and Skehan (1996), which will thus make comparisons possible.

The two tasks differed in a number of aspects. The instruction task tapped familiar information, was very structured, and required the use of the present or future tense. The exposition task, on the other hand, was based on unfamiliar information, was unstructured, and required the use of the past tense. It can thus be argued that the exposition task was cognitively more difficult than the instruction task.

Planning Condition

Planning was operationalized in terms of planning time and represented at four levels between subjects, resulting in four groups. Subjects in the control group had no planning time available. In the three experimental groups, subjects had 1 minute, 5 minutes, and 10 minutes planning time, respectively. In order to ensure that the time provided was indeed used for planning, subjects in the three planning conditions were asked to make written notes.

Subjects and Setting

Subjects were 31 members of two groups of university students studying German at the intermediate level at a university in London. The two groups (altogether 44 students) followed the same German-language program but were taught by different teachers. Proficiency levels of the students as measured by exam results varied, with overall exam results of the previous semester ranging from 42 to 71 out of 100. The only grade available which reflected oral proficiency to some degree was a mark

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

given on a prepared joined oral presentation, which formed part of the overall grade and ranged from 38 to 76. Although all the students, for pedagogical and practical reasons, were required to do the two experimental tasks, several students were excluded from analysis in advance on the basis of their exam results in order to minimize group variability. Three students were well above average with exam grades over 65 and oral presentation marks over 70, and 5 students were well below average with overall exam results under 45 and oral presentation marks under 55.

The remaining 36 students were randomly distributed among the four groups. Care was taken, however, to ensure an even distribution with respect to first language and German learning history. Altogether, there were 16 English native speakers and 20 with other first languages (5 Spanish, 4 French, 3 Russian, 3 Italian, 2 Greek, 1 Finnish, 1 Polish, and 1 Persian). Students also differed in their German learning history; 19 had studied German at school and had obtained an A-level (university entrance requirement) or equivalent, and 17 had done a one-year intensive course at a university before they entered the present course. Of the 36 students, only 31 performed both tasks and were subsequently included in the analyses.

Subjects were tested in groups in a language laboratory during scheduled tuition hours. The subjects had used the language laboratory individually previously but not as a group. The experimenter was introduced to them by their regular tutor as a research student needing their cooperation in testing various tasks for a new German program. Subjects were told that, in return, this would provide them with a good opportunity to speak and practice their German. The particular area of interest was not mentioned and subjects were asked not to talk with fellow students about the tasks they had performed.

Design

Each subject in each of the four groups performed two different tasks over a period of 2 weeks. Order of task was counterbalanced across subjects in each group to control for practice effects.

Subjects in the unplanned condition received the task sheet and were asked to begin as soon as they had read through the task, for which they were given 15 seconds. In the three planned conditions, subjects received the task sheet, which also included an instruction according to which they had 1, 5, or 10 minutes to plan their speech in terms of what to say and how to say it. Because it was not part of the study to test subjects' comprehension, the task itself was described in German and English and the actual instruction was provided in English. This was to ensure that all subjects understood both task and instruction (see Appendix for example task sheets for the 10-minute planning condition).

Subjects were asked to make written notes in German on a separate sheet of paper to make sure they engaged in planning. Subjects were not allowed to write out everything in detail and were reminded that their oral production would be made without the notes. Before starting to speak, subjects had to hand over their note sheets but were allowed to keep the task sheet until they had completed the task.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

Analysis

After transcribing the recorded speech samples, various analyses were carried out to determine the values for each dependent variable, each subject, and each task. What follows is a detailed description of how the speech samples were analyzed.

Fluency Measures. In this study, fluency was measured by number of pauses, total pausing time, mean length of run, and speech rate. The length of all responses was measured using a stop watch and rounded off to whole seconds. The location of potential pauses of 1 second or longer was then determined, the length of each potential pause was measured twice with the help of a stop watch, the mean was taken, and the location and length of all pauses of 1 second or more were indicated in the transcripts. No distinction was made between unfilled pauses and pauses that included fillers such as *hm, um*, and *uh*. Number of pauses was calculated by counting the number of pauses of 1 second or more that occurred in the first 3 minutes of speech (most subjects did not speak for much longer). Total pausing time was calculated by adding up all the pauses and expressing the total as percentage of the total time used for uttering the entire text. Mean length of run was calculated by counting the syllables between pauses and calculating the mean number of syllables.

The next step concerned the identification of all syllables, words, and phrases that were subsequently repeated with or without modifications or abandoned before completion. This was followed by a count of all syllables uttered and a syllable count excluding all syllables that were repeated. Speech rate (A) was calculated by dividing the number of syllables in a given speech sample by the time taken to produce them (measured in seconds) and multiplying the result by 60. Speech rate (B) was calculated the same way as speech rate (A), but all syllables, words, or phrases that were subsequently repeated, reformulated, or replaced were excluded from the count.

Complexity Measures. The transcribed speech data was divided into c-units, T-units, and dependent clauses. Street (1971, cited in Larsen-Freeman, 1978) defines T-units as the shortest possible units that grammatically allow sentence punctuation. The T-unit can be described as one main clause plus whatever subordinate clauses, phrases, and words happen to be attached to or embedded within it. Dependent clauses include finite as well as nonfinite clauses (i.e., subordinate clauses, coordinate clauses with subject deletion, coordinate clauses with subject and finite verb deletion, and infinitive constructions, excluding infinitives used with modal verbs). Following Brock (1986), all of these were taken to signal an underlying s-node. On the basis of this, number of subordinate clauses per T-unit and number of s-nodes per T-unit were determined.

Foster and Skehan (1996) argued that the definition of the T-unit excludes ellipsis and is therefore often unsatisfactory when dealing with spoken interaction. They suggest using the c-unit (communication unit) as a more sensitive measure. Loban (1963, pp. 6–7) defined c-units as "grammatical independent predications or . . . answers to questions which lack only the repetition of the question elements to satisfy the criterion of independent predication." In this study, frequent formulaic utter-

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

ances that do not contain any finite verb, mainly in respect to address and leave-taking such as *hello, good-bye, till tomorrow*, and *sorry*, were identified as c-units but not as T-units. All T-units were also marked as c-units. A word count for each c-unit resulted in the third complexity measure: length of c-units (i.e., the average number of words per c-unit).

Accuracy Measures. The percentage of error-free clauses and the number of errors per 100 words were used as general measures of accuracy. In order to test the possibility that planning may affect various aspects of the language differently with respect to accuracy, the analysis also includes two specific measures: word order and lexical choice error. It is assumed that subjects will use planning time to retrieve and activate lexical items needed to perform the task, which would result in fewer lexical choice errors.

Error analysis to establish number of errors included all errors in syntax, morphology, and lexical choice, but not intonation or pronunciation. An incorrect inflectional ending, no matter whether the mistake was due to, for example, wrong gender, wrong case, or both, was counted as one error. Errors that were repeated (i.e., forms based on wrong assumptions but consistent) were counted only once. Reference errors, such as switches between *du* and *Sie* (informal and formal *you*) were counted as errors every time a shift occurred.

With respect to word order, three word-order rules of German regarding the position of the verbs were investigated. They are:

Separation (*SEP*). If a main clause contains a finite and a nonfinite verb, the finite verb is in clause-second position and the nonfinite verb is in clause-final position.

Inversion (*INV*). If a main clause begins with anything other than the subject, the order of the subject and the finite verb is inverted.

Verb final (*VF*). In a subordinate clause, the finite verb is placed at the end of the clause.

Obligatory contexts for the three word-order rules were identified and the produced word order was scored as correct or deviant.

By adapting error categorizations of Grauberg (1971) and Rogers (1984), lexical errors were defined as errors in lexical choice affecting words, phrases, or collocations. These included completely unintelligible words, English words, or phrases unacceptable in German (e.g., *alarm clock* or *yesterday*, but not proper names such as *Central Line, Ealing Shopping Centre*), lexical items that were confused due to related meaning or due to misleading equivalence of English and German words (e.g., *gehen–fahren–kommen*, "to go on foot"–"to go by car, bus, etc."–"to come"; *geben–bringen*, "to give"–"to bring"; *fragen–bitten*, "to ask about"–"to ask for"; *als–wenn*, "at the time when"–"whenever"), verbs with the prefix omitted or wrong (e.g., *ich werde Sie von der U-Bahn holen* instead of *abholen*, "I will get you from the tube station" instead of *pick up*), incorrect prepositions (e.g., *du gehst nach meine Universität* instead of *du gehst zu meiner Universität*, "you go to my university"), missing prepositions in phrasal verbs (e.g., *ich warte dich* instead of *ich warte auf dich*, "I wait you" instead of "I wait for you"), and contextually inappropriate expression or meaning (e.g., *ich kann dich nicht am Flughafen treffen* instead of *ich kann*

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

dich nicht vom Flughafen abholen, "I cannot encounter you at the airport" instead of "I cannot pick you up from the airport"; *hallo, ich bin Susanna* instead of *hallo, hier ist Susanna*, "hello, I am Susanna" instead of "hello, this is Susanna"). Lexical errors were identified by the author and a second rater with an interrater reliability score of 74%. Discrepancies on the remaining 26% were resolved by discussion.

Density Measures. Density of speech was measured here by weighted lexical density. In order to measure lexical density, it was necessary to divide linguistic items into grammatical and lexical items. A linguistic item may consist of more than one word, which means that, for example, phrasal verbs or idioms were counted as one linguistic item. In addition, different words may realize one lexeme (e.g., *different* and *difference*). The classification of items used was based on the framework described by O'Loughlin (1995). Lexical items were further divided into high-frequency and low-frequency lexical items. High-frequency items were those that occurred more than once in an individual speech sample. High-frequency lexical items were given half the weight of low-frequency lexical items, and the weighted number of lexical items as percentage of total number of items was then calculated.

RESULTS

The study followed a 4 × 2 research design with two independent variables: planning time and task type. The planning variable was a between-subjects factor represented at four levels. The task variable was a within-subjects factor with two levels. A number of statistical analyses were performed: (a) a series of *t* tests to reveal any possible practice effects, (b) a series of ANOVAs to test for differences between planning conditions, (c) a MANOVA to test for any interaction between planning and task type, (d) a factor analysis to see how many factors underlie the set of 13 dependent variables, and (e) a MANOVA with only four dependent variables, representing the four areas of speech performance under consideration, to test for the overall effect of planning on speech performance. This MANOVA was performed last because the four dependent variables were chosen on the basis of the results of the ANOVA tests and the factor analysis. All statistical analyses were performed using the SPSS/PC+ package.

Practice Effects

The first analysis performed concerned the possible effect of task order. Although order of task was counterbalanced within each group, it is possible that no matter which task was performed first, subjects' performance would improve on the second task as a result of practice. A series of *t* tests were carried out on each dependent variable for each of the two tasks. The results are given in Table 1, which gives the mean of each dependent variable for subjects who did the task first (M1) and subjects who did the task second (M2) as well as *t* values and probabilities.

Results show that statistical significance for the effect of order of task on speech performance was not attained. However, there seems to be a trend indicating that

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

Table 1. Results from the *t* tests on possible practice effects

	Instruction Task				Exposition Task			
	*M*1	*M*2	*t*	*p*	*M*1	*M*2	*t*	*p*
Speech rate A	81.1	100.5	1.82	.08	90.7	101.9	0.94	.35
Speech rate B	72.7	90.0	1.65	.11	80.5	93.0	1.06	.29
Mean length of run	9.8	13.2	1.49	.15	10.9	13.5	0.73	.47
Total pausing time	39.4	32.4	1.38	.18	36.4	37.4	0.17	.87
Number of pauses	26.6	25.6	0.37	.72	25.2	26.9	0.65	.52
Words per c-unit	7.4	7.3	0.27	.79	7.1	7.1	0.02	.98
Subclauses/T-unit	19.0	21.9	0.55	.59	30.1	31.6	0.23	.82
S-nodes/T-unit	23.5	27.9	0.71	.48	48.9	45.6	0.46	.65
Errors/100 words	19.6	19.7	0.02	.98	22.8	25.8	0.71	.49
Error free clauses	42.5	47.5	0.72	.48	41.3	37.4	0.54	.59
Word order errors	24.2	27.3	0.41	.68	23.2	20.0	0.39	.70
Lexical errors	7.5	5.2	1.63	.11	6.3	7.3	0.74	.47
Weighted lexical density	26.3	28.5	1.94	.06	32.0	32.5	0.29	.77

Note. For each task and each dependent variable, two means are given depending on whether the task was done first or second.

practice had an influence on the level of performance achieved. This was shown most clearly with subjects who performed the instruction task after the exposition task. They showed better performance on the instruction task, in particular on fluency measures and lexical density, compared to subjects who performed the instruction task first. It is interesting to note that Foster and Skehan (1996) also observed slight practice effects, which in their case were evidenced for accuracy. Because sample sizes in both studies were quite small, further research is clearly needed to determine whether practice can have a significant effect on speech performance and of what nature that effect would be.

Planning Effects

The first statistical test on the planning variable concerned its effect on the dependent variables for the two tasks. In order to determine for which measures differences reached significance, one-way ANOVAs were carried out on each dependent variable. When significance was reached, the Student-Newman-Keuls procedure was used to establish between which groups differences were significant. Results for each task are given in Tables 2 and 3, with the *F*-value, the probability, means for the four groups, and an indication of where differences reached significance.

Hypothesis 1 predicted that planning would result in more fluent speech. Fluency was measured by two types of speech rate, number of pauses, mean length of run, and total pausing time. Results show that differences between planners and nonplanners were clearest on the measures of speech rate and total pausing time. Number of pauses and mean length of run showed similar patterns, with significant differences only between nonplanners and 10-minute planners for the structured

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

Table 2. Results from the ANOVA on the planning variable for the instruction task

			Planning Time				Location of Significant Difference
	F	*p*	0	1	5	10	
Speech rate A	4.38	.012	63.4	89.8	99.5	109.1	0–5, 10
Speech rate B	5.00	.007	55.1	76.8	89.6	102.3	0–5, 10
Mean length of run	6.52	.002	7.2	10.1	10.2	18.2	0, 1, 5–10
Total pausing time	6.55	.002	49.0	38.7	31.8	24.9	0–5, 10
Number of pauses	3.81	.021	27.2	28.6	29.2	19.8	0, 1, 5–10
Words per c-unit	0.36	.782	7.0	7.3	7.4	7.7	
Subclauses/T-unit	2.45	.085	17.4	18.0	14.7	31.2	
S-nodes/T-unit	1.57	.220	21.8	21.0	22.6	36.6	
Errors/100 words	3.59	.026	27.6	17.3	19.0	14.3	0–10
Error-free clauses	2.18	.114	32.0	51.0	43.9	53.5	
Word order errors	2.04	.132	39.4	25.1	22.8	15.5	
Lexical errors	1.91	.152	8.4	7.3	6.1	3.9	
Weighted lexical density	2.01	.136	25.5	27.0	29.2	27.7	

Note. In the column "Location of Significant Difference," an entry such as "0–5, 10" means that significant differences were found between nonplanners and 5-minute planners as well as between nonplanners and 10-minute planners.

Table 3. Results from the ANOVA on the planning variable for the exposition task

			Planning Time				Location of Significant Difference
	F	*p*	0	1	5	10	
Speech rate A	5.08	.006	63.8	99.9	110.7	110.4	0–1, 5, 10
Speech rate B	5.03	.007	55.3	86.9	101.6	102.4	0–1, 5, 10
Mean length of run	1.88	.156	6.9	10.4	13.8	17.4	
Total pausing time	8.75	.000	55.2	34.1	28.0	29.7	0–1, 5, 10
Number of pauses	1.45	.251	28.2	29.5	23.9	23.0	
Words per c-unit	1.10	.367	7.7	6.3	6.8	7.5	
Subclauses/T-unit	0.51	.682	32.1	23.9	32.4	34.5	
S-nodes/T-unit	1.50	.238	49.8	41.0	40.2	59.0	
Errors/100 words	3.61	.026	33.4	16.3	23.9	22.5	0–1
Error-free clauses	2.14	.118	27.4	51.4	38.5	41.8	
Word order errors	0.59	.628	27.5	15.5	16.7	26.2	
Lexical errors	0.85	.479	8.1	5.4	6.3	7.3	
Weighted lexical density	4.26	.014	29.4	31.1	31.6	36.8	0, 1, 5–10

Note. In the column "Location of Significant Difference," an entry such as "0–5, 10" means that significant differences were found between nonplanners and 5-minute planners as well as between nonplanners and 10-minute planners.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

task, but no significant differences for the unstructured task. Overall, it can be concluded that nonplanners clearly produced less fluent speech than planners and that Hypothesis 1 is confirmed.

Before discussing Hypothesis 2, let us consider Hypothesis 5 and its application to fluency. Hypothesis 5 predicted that the effect of planning would increase in line with the amount of planning time available; that is, 1-minute planners would be more fluent than nonplanners but less fluent than 5-minute planners, and 10-minute planners would be the most fluent of all. Although most differences between planning groups did not reach significance, results show that speech rate increased and total pausing time decreased with amount of planning time for the instruction task. The extent of the effect of planning, however, diminished with more planning time. A similar pattern can be observed for the exposition task, in which there is hardly any difference between 5- and 10-minute planners. Mean length of run shows a steady increase across planning conditions, especially for the exposition task. With regard to number of pauses, the clearest effect of planning can be observed for the 10-minute planners in the instruction task and the 5- and 10-minute planners in the exposition task. It can thus be concluded that, with respect to fluency, Hypothesis 5 received broad confirmation.

Hypothesis 2 predicted that planned speech would show greater complexity of language, which was measured by words per c-unit, number of subordinate clauses, and number of s-nodes. There were no significant differences, either between nonplanners and planners or among planners, for words per c-unit. With regard to the other two measures, the speech of 10-minute planners showed more subordinate clauses and more s-nodes in the instruction task than the other three groups. Results for the exposition task are more complex with 1-minute planners producing fewer subordinate clauses than the other groups and 10-minute planners producing more s-nodes than nonplanners, who in turn produced more than 1- and 5-minute planners. None of these differences were significant. Hypothesis 2 thus receives only limited support. Hypothesis 5, predicting an increase in complexity with more planning time, was not confirmed because the clearest differences were between nonplanners and 10-minute planners, with 1- and 5-minute planners showing the same or slightly lower scores than nonplanners.

Hypothesis 3 predicted that planners would produce more accurate speech. With regard to number of errors, this hypothesis was confirmed because nonplanners produced significantly more errors than planners. A similar trend was obtained with error-free clauses, the second general accuracy measure, although differences here did not reach significance. With regard to the two specific measures, word-order errors, and lexical choice errors, differences between nonplanners and planners were in the predicted direction but not statistically significant. Overall, however, Hypothesis 3 can be said to be confirmed. Looking at differences between planning conditions, results are clearest for the instruction task on number of errors and the two specific measures showing an increase of accuracy with amount of planning time. In the exposition task, differences were clearest between nonplanners and 1-minute planners, with 5- and 10-minute planners producing less accurate language than 1-minute planners. Hypothesis 5 regarding accuracy thus receives no support.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

Table 4. Results from the ANOVA on the planning variable for the instruction and exposition tasks combined

			Planning Time				Location of Significant Difference
	F	*p*	0	1	5	10	
Speech rate A	5.75	.004	63.6	94.9	105.1	109.8	0–1, 5, 10
Speech rate B	6.39	.002	55.2	81.8	95.6	102.4	0–1, 5, 10
Mean length of run	3.90	.019	7.0	10.2	12.0	17.8	0–10
Total pausing time	10.94	.0001	52.1	36.4	29.9	27.3	0–1, 5, 10
Number of pauses	2.77	.061	27.7	29.0	26.6	21.4	
Words per c-unit	0.59	.630	7.3	6.8	7.1	7.6	
Subclauses/T-unit	1.42	.260	24.7	21.0	23.5	33.3	
S-nodes/T-unit	2.97	.050	35.8	31.0	31.4	48.6	5–10
Errors/100 words	3.97	.018	30.5	16.8	21.4	18.4	0–1, 5, 10
Error free clauses	2.94	.051	29.7	51.2	41.2	47.7	
Word order errors	1.24	.315	33.5	20.3	19.7	20.8	
Lexical errors	0.97	.420	8.3	6.3	6.2	5.6	
Weighted lexical density	3.56	.027	27.5	29.0	30.4	32.3	0–10

Note. In the column "Location of Significant Difference," an entry such as "0–5, 10" means that significant differences were found between nonplanners and 5-minute planners as well as between nonplanners and 10-minute planners.

Interestingly, the pattern for error-free clauses differs from the other three accuracy measures but also confirms that differences were clearest between nonplanners and 1-minute planners.

Hypothesis 4 predicted that planned speech would show greater density as measured by weighted lexical density. Results were in the predicted direction but significant for the exposition task, where the greatest difference was between nonplanners and 10-minute planners, with the other two planning groups intermediate. Hypothesis 4 and Hypothesis 5 thus receive some support with regard to density of language.

Composite Planning Effects

Collapsing the results for the two different tasks, Table 4 shows the composite results for effect of planning. The performance of nonplanners and planners differs significantly on four of the five fluency measures. Speech rate and total pausing time change across planning conditions, but the size of improvement decreases with amount of planning time. With regard to mean length of run, only 10-minute planners differ significantly from nonplanners, with 1- and 5-minute planners intermediate. Similarly, the number of pauses is smaller for 10-minute planners in comparison with the other three groups, but the difference is not statistically significant.

Of the three complexity measures, only number of s-nodes per T-unit shows any significant difference, which is between 5- and 10-minute planners, with 1-minute planners having similar and nonplanners slightly higher numbers of s-nodes than 5-minute planners.

With regard to accuracy, differences in number of errors reach significance be-

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

tween nonplanners and all planners, but no differences were found across the planning groups. The second global measure and the two specific measures show a similar pattern, but differences between nonplanners and planners are not significant.

Weighted lexical density increases with amount of planning time, although the only difference that reached significance was between nonplanners and 10-minute planners.

Interaction Between Planning and Task Type

In order to test Hypothesis 6, which predicted that the effects of planning would be greater for the cognitively more difficult task, a MANOVA including all 13 measures was carried out. Analysis revealed no significant interaction effect between planning and task type (Pillais = 1.65, $F = 1.51$, $p = .09$). However, univariate F tests showed that significance was almost reached with regard to weighted lexical density ($F = 2.82$, $p = .059$), lexical error ($F = 2.42$, $p = .089$), and words per c-unit ($F = 2.33$, $p = .098$). From Tables 2 and 3, we can see that the planning effect on weighted lexical density and words per c-unit was greater in the exposition task, but for lexical error the planning effect was greater in the instruction task.

The MANOVA test also revealed a significant effect of task type on speech performance (Pillais = 0.87, $F = 6.09$, $p = .001$), with differences being concentrated in number of subclauses (instruction task: $M = 20.3$, exposition task: $M = 30.8$), number of s-nodes (instruction task: $M = 25.7$, exposition task: $M = 47.3$), number of errors (instruction task: $M = 19.6$, exposition task: $M = 24.3$), and weighted lexical density (instruction task: $M = 27.4$, exposition task: $M = 32.3$). The latter is, however, an artifact of analysis because the instruction task required the use of many place names and these proper names were not included in the count.

Underlying Constructs

A number of variables were used to measure fluency, accuracy, and complexity. In order to uncover whether the measures in these three sets of variables are related to one another and whether they really represent three distinct factors, the data were subjected to a factor analysis. Weighted lexical density was also included in the analysis to see whether this variable measures a distinct aspect of language performance. The results obtained are shown in Table 5, which contains factor loadings of only .40 and above, and communality scores, which indicate the amount of variance that is explained by the factors.

Factor 1 is made up of four high loadings on the two speech rate measures, mean length of run, and total pausing time. It would appear that general fluency affects all these different measures, bringing together general rate of speech as well as amount of speech between pauses. Number of pauses seems to be slightly less central to this. However, these five loadings confirm the existence of a general fluency factor. More interesting is the way weighted lexical density also loads on this factor. This means

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

Table 5. Results from the factor analysis

Measures	Factor 1	Factor 2	Factor 3	Communality
Speech rate A	.82	−.43		.85
Speech rate B	.85	−.42		.89
Mean length of run	.84			.84
Total pausing time	−.88			.88
Number of pauses	−.64			.45
Words per c-unit			.88	.84
Subclauses/T-unit			.96	.92
S-nodes/T-unit			.93	.87
Errors/100 words		.87		.91
Error-free clauses		−.82		.82
Word order errors		.77		.66
Lexical errors		.84		.81
Weighted lexical density	.62			.51

Note. Factor loadings of only .40 and above are included.

that those subjects whose performance was more fluent also used a higher proportion and a greater range of lexical words.

The three error measures and number of error-free clauses load highly on the second factor, confirming that all four variables are related to one underlying construct that can be taken to represent speech accuracy. Interestingly, the two speech rate measures also load on this factor, although loadings are quite low. A third independent factor underlies the data, which consists of words per c-unit, number of subordinate clauses, and number of s-nodes. All three variables load very highly on this factor and can be taken to define that factor as signifying speech complexity.

Overall Effect of Planning on Speech Performance

So far, tests of group differences were performed on each dependent variable separately. However, there is some degree of interrelationship between the dependent variables, as shown by the factor analysis, that cannot be ignored. Therefore, a MANOVA, which considers the dependent variables simultaneously, was carried out. Because the four areas of speech performance (i.e., fluency, complexity, accuracy, and density) were measured by different numbers of variables, it did not seem appropriate to include all variables in the statistical test. On the basis of the results of the ANOVA tests and the factor analysis, speech rate (A), number of s-nodes, and number of errors, which best reflected differences between groups and had high factor loadings, were chosen in addition to weighted lexical density to represent the four areas of speech performance. Analysis revealed a highly significant difference between planning conditions on the set of the four dependent variables (Pillais = 0.92, $F = 2.75$, $p = .004$), showing a significant overall effect of planning on speech performance.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

DISCUSSION

The present study focused on the effect of planning on L2 learners' speech performance. It was assumed that planning time is used by L2 learners to prepare cognitively and linguistically; that is, to decide on what meaning they want to convey and to search for and activate the linguistic resources best suited to express the intended meaning. This pretask planning was assumed to ease the processing load during task completion by reducing the amount of online planning necessary, which in turn would enable the speaker to produce more fluent, complex, and accurate, as well as more dense, language.

Comparing the performance of nonplanners and 10-minute planners, results show that 10-minute planners produced significantly more fluent and more accurate speech with higher lexical density. Their speech was also more complex, although the difference between nonplanners and 10-minute planners did not reach significance. The results regarding fluency compare with the findings of Foster and Skehan (1996). With regard to number of s-nodes and subordinate clauses, the findings of this study replicate the results reported by Foster and Skehan (1996) and Crookes (1989). The results for accuracy, whereas previous studies obtained conflicting evidence, corroborate those found by Foster and Skehan (1996) and Wigglesworth (in press). Of particular interest are the results for weighted lexical density, which provide evidence to confirm the assumption that the degree to which discourse is planned is a source of difference in spoken language in this regard.

The central focus of this study was the influence of amount of planning time on the size of the planning effect. It was hypothesized that any effect of planning would increase in line with amount of planning time made available. With regard to fluency, it is interesting that different measures show different patterns (see Figures 1–4).

Speech rate increases and total pausing time decreases with amount of planning time, but the improvement diminishes with more planning time. Mean length of run, on the other hand, increases with no sign of diminishing, whereas number of pauses reduces only after 5 minutes planning time. As with mean length of run, weighted lexical density also steadily increases across planning conditions. The clearest difference for complexity is between 10-minute planners and the other three groups, whereas for accuracy, the greatest improvement was recorded for the 1-minute planners, with the 5- and 10-minute planners showing the same or slightly lower levels of improvement.

We can conclude then that more planning time results in improved performance overall, but the increase in performance is not the same for the four areas of performance considered here (see Figures 5–8), which means that different things are going on at different times. A comparison of nonplanners and 1-minute planners showed that performance improved with regard to fluency, density, and accuracy. The performance of 1- and 5-minute planners differs only with respect to fluency and density, whereas 10-minute planners differ from 5-minute planners in fluency and density of speech produced as well as complexity of language used. In other

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

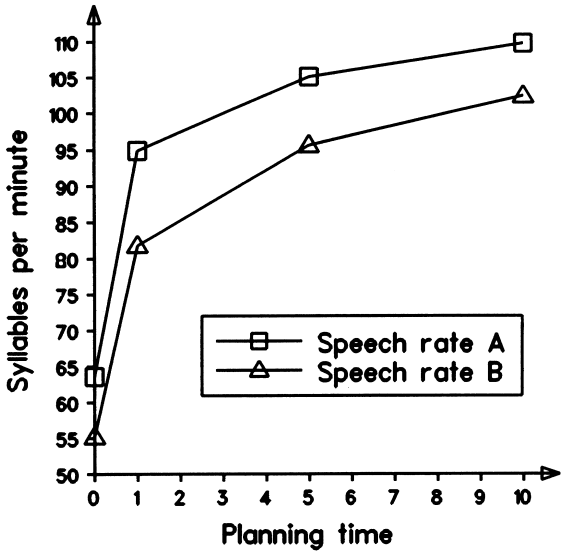


Figure 1. Influence of amount of planning time on speech rate.

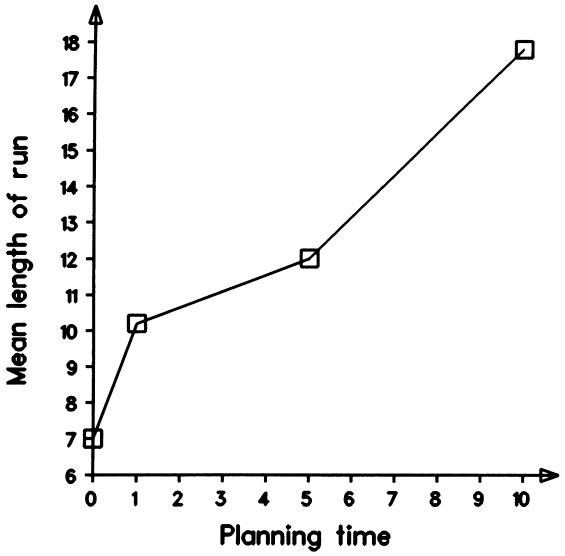


Figure 2. Influence of amount of planning time on mean length of run.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

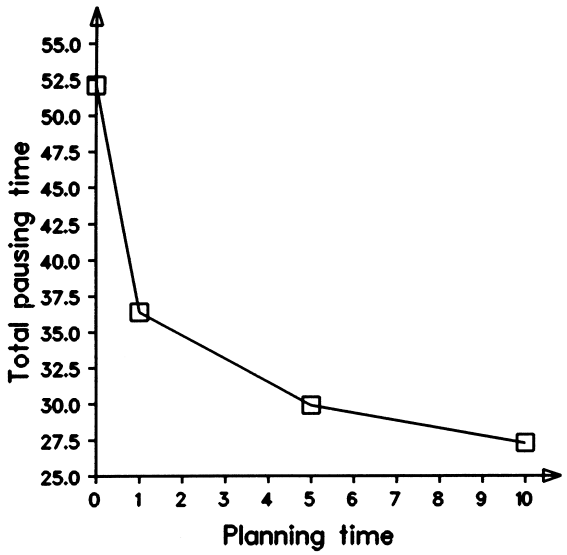


Figure 3. Influence of amount of planning time on total pausing time.

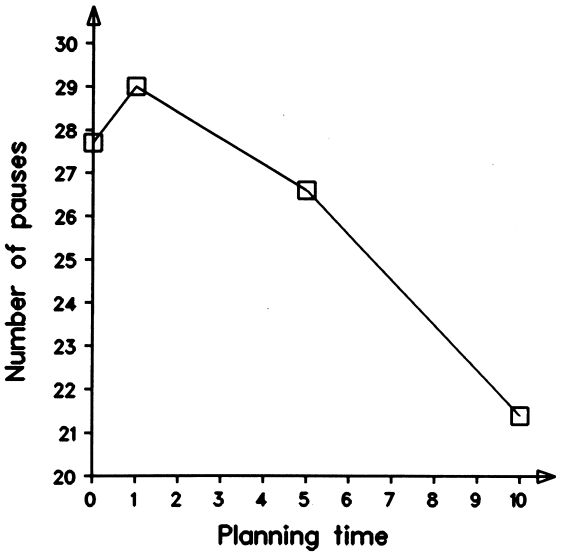


Figure 4. Influence of amount of planning time on number of pauses.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

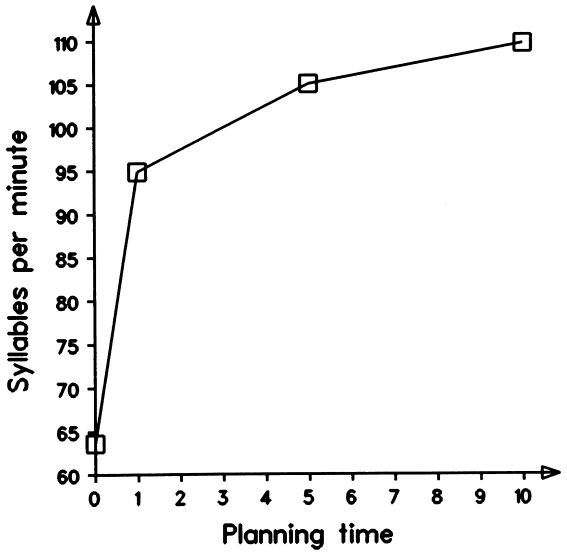


Figure 5. Influence of amount of planning time on fluency (speech rate A).

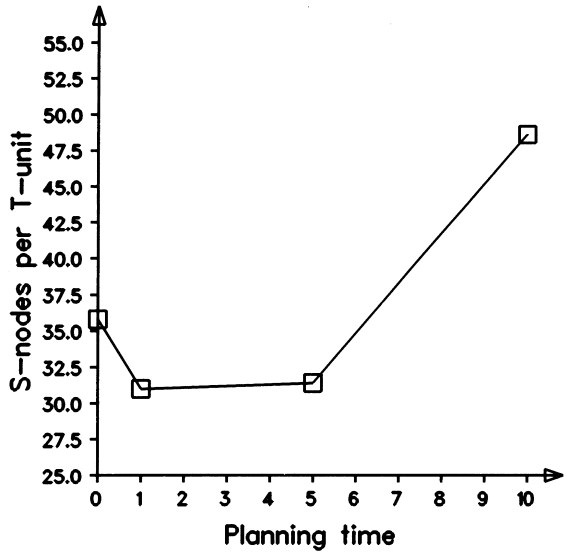


Figure 6. Influence of amount of planning time on complexity (number of s-nodes).

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

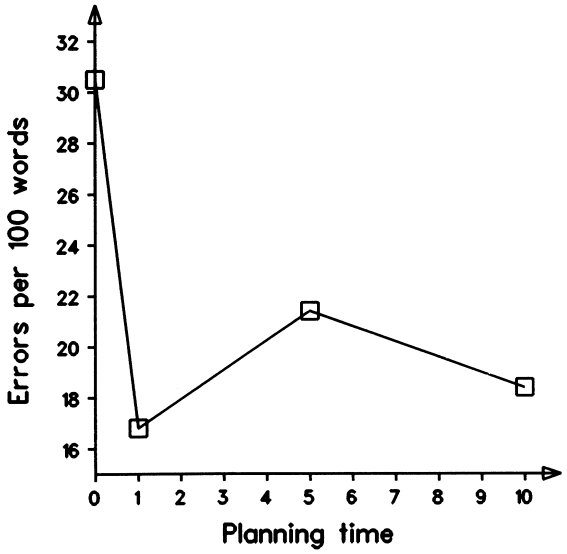


Figure 7. Influence of amount of planning time on accuracy (number of errors).

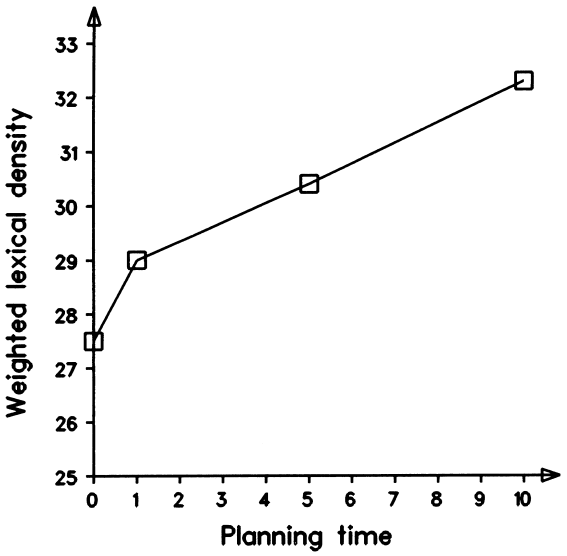


Figure 8. Influence of amount of planning time on density (weighted lexical density).

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

words, when subjects had only 1 minute to plan, priority was given to accuracy. Only when subjects had 10 minutes planning time did they attempt to produce more complex language.

These results seem to provide support for the notion of an existing competition for attentional resources between the goals of accuracy and complexity. When provided with planning time, the first priority was accuracy, which was achieved in the first minute and did not improve with more planning time and in the case of the exposition task actually declined again with 5- and 10-minute planners. Greater complexity, on the other hand, was only achieved by the 10-minute planners, who used the greater amount of planning time for allocating attentional resources to complexity to the exclusion of further improvements in accuracy. The conclusion that can be drawn is that planning leads to performance improvements, but any gains in accuracy and complexity are not simultaneously achieved. This is because attentional resources are limited and are therefore allocated either towards achieving accuracy or complexity at any one time.

With regard to differences between the two tasks, it was argued that the exposition task should be more difficult than the instruction task because it relied on less familiar information, was less structured, and required for its completion the use of more difficult linguistic material. It was further argued that performance on the exposition task should derive greater benefit from pretask speech planning. As the two tasks employed in this study share similarities with the personal information exchange and narrative tasks in Foster and Skehan's (1996) study, a comparison of results could be revealing. Foster and Skehan reported an interaction between planning condition and task type, with the personal task producing higher fluency gains under planning conditions and the narrative task higher complexity gains. Although the present study did not reveal any significant interaction effects between planning and task type, the effect of planning was greater for lexical errors on the instruction task and for length of c-unit on the exposition task.

Foster and Skehan (1996) also found that, under the unplanned condition, performance on the personal task showed greater fluency than on the narrative task, which in contrast had higher complexity scores than the personal task. In this study, the two tasks did not produce different results for fluency. However, the instruction task, which compares with Foster and Skehan's personal task, is characterized by greater accuracy (number of errors), whereas the exposition task produced higher complexity scores (number of subclauses and s-nodes) than the instruction task. What emerges is that when tasks have a clear structure and do not require subjects to express complex ideas, priority is given to achieving fluency (Foster & Skehan, 1996) or accuracy (this study). On the other hand, when more complex language needs to be used, limited attentional resources force subjects to concentrate on complexity to the disadvantage of fluency and accuracy. It seems then that task properties, at least partly, determine the level of fluency, accuracy, and complexity produced, as well as where attentional resources are allocated when planning time is available.

Turning finally to the question of which are the optimal measures to select for quantifying fluency, complexity, and accuracy, it was found that the results of this

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

study present an interesting picture. Previous research into planning differed mainly with regard to the measures of accuracy adopted. Crookes (1989) and Wigglesworth (in press) used specific measures, such as plural *-s*, the article, and verb morphology, whereas Foster and Skehan (1996) supported more global measures and adopted error-free clauses, arguing that more general measures are better suited to detect differences because they take into account all contributory influences on error and correctness. In this study, both global and specific measures were included in the analysis. Because the use of error-free clauses does not differentiate between clauses with one or more errors, a second global measure—number of errors per 100 words—was selected to measure accuracy of speech, which was believed to be even more sensitive to existing differences. The results support this assumption. Although differences in error-free clauses were in the predicted direction, only differences in number of errors were statistically significant. With regard to the two specific measures used (i.e., word order and lexical errors), differences again were in the predicted direction but not statistically significant. From these results, it can be concluded that more global measures are indeed more suitable for detecting any differences and that number of errors seems to be an even more sensitive measure than error-free clauses, at least for learners of German at the early-intermediate level. However, global measures do not show whether all types of errors are equally affected by planning or only some of them. Only a detailed classification and analysis would reveal any differences in the effect of planning on different language forms.

The results for complexity also support the argument for the use of more global measures. Here the number of subordinate clauses (as used by Crookes, 1989, and Wigglesworth, in press) was contrasted with number of s-nodes (a similar measure to number of clauses as used by Foster & Skehan, 1996). S-nodes include subordinate clauses as well as nonfinite clauses and could therefore capture a greater deal of variance. This assumption was borne out by the results of this study, with the number of s-nodes per T-unit as the only complexity measure to show a significant effect for planning.

As regards fluency, the different measures seem to fall into two broad groups. In one group are number of pauses and mean length of run, which are based on the number of times subjects are forced to pause, not taking into account the length of pauses. The results with regard to these two measures show similar patterns. The second group comprises the two speech measures and total pausing time, which center on the amount of time spent pausing as opposed to time spent speaking, no matter whether pausing time was spread over just a few or many pauses. These measures seem to capture the effect of planning and differences between planning groups most clearly.

CONCLUSION

The results of this study provide further evidence of the influence of planning on the speech performance of L2 learners, demonstrating that results can be generalized to learners of languages other than English. With regard to the effect of amount of planning time, this study corroborates results by Wigglesworth (in press), showing

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

that as little as 1 minute of planning time can lead to clear changes in L2 learners' speech performance, and this has to be taken into account, especially in language-testing situations. On the other hand, it confirms Foster and Skehan's (1996) claim that speech planning as a pretask activity is a useful pedagogical tool to manipulate task conditions and influence task outcome.

The findings of this study resulting from a systematic investigation of the influence of amount of planning time on L2 learners' speech performance are an important precondition for subsequent studies in this area. Most of the L2 planning research so far has operationalized planning in terms of providing subjects with planning time and making them take notes. How to approach the planning task was very much left to the subjects themselves. It could therefore be argued that there were great individual differences between subjects as to the activities they engaged in and that most of them did not use the time provided in the most effective manner, especially because many subjects were faced with such a speech planning task for the first time. Therefore, it seems useful to investigate methods to make L2 learners more effective planners, such as with the help of instruction, continued practice, or consciousness-raising activities like verbalization. Closely linked to that is research into methods of how minimal planning time can be used most productively and how subjects' attention can be channelled intentionally into improving either fluency, accuracy, or complexity of speech, or all of them equally.

(*Received 18 February 1996*)

REFERENCES

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition, 11*, 17–34.

Brock, C. (1986). The effects of referential questions on ESL classroom discourse. *TESOL Quarterly, 20*, 47–59.

Clark, H., & Clark, E. (1977). *Psychology and language. An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

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

Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon Press.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of past tense. *Studies in Second Language Acquisition, 9*, 1–20.

Færch, C., & Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Færch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 20–60). London: Longman.

Færch, C., & Kasper, G. (1986). Cognitive dimensions of language transfer. In E. Kellerman & M. Sharwood Smith (Eds.), *Cross-linguistic influence in second language acquisition* (pp. 49–65). Oxford: Pergamon Press.

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

Grauberg, W. (1971). An error analysis in German of first-year university students. In G. Perren & J. Trim (Eds.), *Applications of linguistics* (pp. 257–263). New York: Cambridge University Press.

Halliday, M. (1985). *Spoken and written language*. Melbourne, Australia: Deakin University Press.

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly, 12*, 439–448.

Loban, W. (1963). *The language of elementary school children* (Research Rep. No. 1). Champaign, IL: National Council of Teachers of English.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. Dechert, D. Möle, & M. Raupach (Eds.), *Second language production* (pp. 26–49). Tübingen, Germany: Gunter Narr.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing, 12*, 217–237.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 263–270). The Hague: Mouton.

Rogers, M. (1984). On major types of written error in advanced students of German. *International Review of Applied Linguistics, 22*, 1–39.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*, 38–62.

Towell, R. (1987). Approaches to the analysis of the oral language development of the advanced language learner. In J. Coleman & R. Towell (Eds.), *The advanced language learner* (pp. 157–182). London: CILTR.

Ure, J. (1971). Lexical density and register differentiation. In G. Perren & J. Trim (Eds.), *Applications of linguistics* (pp. 443–452). New York: Cambridge University Press.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*, 85–106.

Zora, S., & Johns-Lewis, C. (1989). Lexical density in interview and conversation. *York Papers in Linguistics, 14*, 89–100.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press

APPENDIX

TASK SHEET FOR THE INSTRUCTION TASK

Aufgabe

Ein deutscher Freund von Ihnen kommt morgen nach London, um Sie zu besuchen. Sein Flugzeug landet um 11 Uhr in Heathrow. Da Sie zu der Zeit in der Universität sind, können Sie ihn nicht vom Flughafen abholen.

Rufen Sie Ihren Freund an und hinterlassen Sie eine Nachricht auf seinem Anrufbeantworter. Erklären Sie ihm, wie er von Heathrow zu Grove House kommt und wann und wo Sie ihn dort treffen werden.

(A German friend of yours will arrive in London tomorrow to visit you. His plane will land in Heathrow at 11 a.m. Since you will have to be at the university at that time, you will not be able to meet him at the airport.

Call your friend and leave a message on his answerphone. Tell him how he will get from Heathrow to Grove House, and when and where you will meet him there.)

Instruction

Before you leave your message, you have 10 minutes to plan what to say and how to say it. Please make written notes in German, but do not write out everything in detail. You then have to talk without your notes.

TASK SHEET FOR THE EXPOSITION TASK

Aufgabe

Sie hatten sich für gestern vormittag mit zwei deutschen Mitstudenten verabredet, um ein Gruppenprojekt für die Universität zu besprechen. Da es Sonntag war, hatten Sie aber plötzlich keine Lust mehr und sind lieber im Bett geblieben. Jetzt müssen Sie sich entschuldigen und brauchen eine gute Ausrede.

Erfinden Sie eine Geschichte, warum Sie nicht kommen konnten. In der Geschichte müssen die folgenden Wörter vorkommen: *Haus, bringen, rot, Geld, kennen, Regenschirm.*

Rufen Sie Ihre Mitstudenten an und hinterlassen Sie Ihre Nachricht auf dem Anrufbeantworter.

(You had arranged to meet yesterday morning with two German fellow students to talk about a group project for university. Since it was Sunday, you no longer wanted to go and rather stayed in bed. Now you have to apologize and you need a good excuse.

Make up a story of why you couldn't come. The story must contain the following words: *house, to bring, red, money, to know, umbrella.* Call your fellow students and leave a message on their answerphone.

Instruction

Before you leave your message, you have 10 minutes to plan what to say and how to say it. Please make written notes in German, but do not write out everything in detail. You then have to talk without your notes.

https://doi.org/10.1017/S0272263198001041 Published online by Cambridge University Press