The Effect of Planning on Oral Output

ROBSON, Graham

Introduction

When L1 speakers engage in conversation, the planning and resultant output is a simultaneous event. However, for L2 speakers on the other hand, it is a much different situation entirely, as any teacher of oral communication classes can testify. Students' speech is often slow and stilted or full of errors, sometimes both. One possible solution to this problem is to give students thinking and planning time before a speaking activity to hopefully make the speech delivery easier, smoother and of a higher quality syntactically. This paper will show how second language acquisition research has used the construct of planning before activities to improve oral output in second language. The introduction starts with a model of planning processes to explain the construct of planning in the mind of a speaker before production, and moves on to a model of language acquisition to connect how planning in the mind can result in spoken output.

The planning process model proposed by Levelt (1989) has been used in the field to explain how speakers use planning before speaking. Levelt defines planning as "…an initial phase in which the speaker decides on a purpose for his next move" (p. 3). The first step in this model is conceptualization, where the message content takes shape. Next, is formulation, as speakers begin to

organise the message into coherent blocks, and lastly, articulation is the physical execution of speech, paying attention to sound and grammar from long-term memory. These three planning levels take place simultaneously in the mind of the speaker, but are also available for returning to the first stage, conceptualization to determine how successful the speech is in meeting the purpose of the speaker. For example, on the one hand, students concerned with the message plan at the conceptual level, and, on the other hand, those who plan at the formulation or articulation level are more concerned with grammar. To put it another way, reaching a balance of message and grammar is not as easy as it sounds as learners have only limited resources available and end up prioritising one aspect of language over another (Van Patten, 1990).

Planning, as I have said, implies a balance in spoken output between focus on forms, or syntactic-based grammar in output, and focus on meaning, in which only the content is of importance when speaking. Obviously, both of these situations in isolation are undesirable to teachers who need to balance forms and meaning. One compromise between the two important features of focus on forms and meaning in speech is referred to as focus on form (Long, 1991). Long describes focus on forms as the need to communicate while focusing on grammar issues. In other words the issue is not sacrificing grammar for meaning, or meaning for grammar.

With the benefit of planning, students can relieve some cognitive burden and allocate language resources they need to push their planned output to the limits of their interlanguage range through a focus on form. At this limit mistakes are more frequent, but output becomes a vehicle for speakers noticing that a need to express a certain form is beyond their current interlanguage ability. Noticing can push the interlanguage forward and create acquisition (Long, 1996; Gass, 1997, p.132), a change in learning, where groups of forms can be accessed quicker from the memory. In other words planning and the resultant focus on form can help students become more proficient by

pushing their interlanguage forward, making acquisition possible and contributing towards determining what aspects of the language become automatized (McLaughlin, 1990), or produced with limited conscious effort.

After establishing planning as an essential element of output, I shall move onto output measurements. In order to measure focus on form in output I will operationalize it into the composite parts. These parts include fluency, or a skill that reflects the primacy of meaning and ability to deal with communication in real time (Schmidt, 1990). After that is complexity, or how elaborate the construction of language is at or above the clause level (Yuan & Ellis, 2003 p. 2); and lastly, accuracy, described as freedom from error and target-like use of language (Foster & Skehan, 1996, p. 304). All of the studies included in this paper focus on these three measurements.

In summary, planning is a process that can help to develop language and determine how speakers prioritise the oral outcomes of fluency, complexity, and accuracy. This paper looks at current research in L2 related to planning in order to better understand the planning process and find an answer to the following question:

How effectively can planning before a spoken activity help students produce higher levels of accuracy, fluency and complexity?

I will now summarize five articles from the field based upon the following: (a) a rationale, (b) subjects, method, hypothesis, results and conclusions, called the study overview and (c) a critical analysis of that research. Finally, the conclusion will tackle current problems and possible ways of improving current research with an eye on the above question.

Article reviews and analysis

Article One - Review

Foster, P. & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 9,*

299-323.

Rational

In this study, a variety of tasks to enhance opportunities for interaction were used to determine their effect on three types of planning.

Study overview

This study focuses on four groups of eight college students (n=32) at pre-intermediate level from a variety of L1 backgrounds, aged between 18 and 30, studying English six hours a week. Two control groups received no planning before attempting the tasks. The two experimental groups were broken down further into two halves, one half getting 10 minutes unguided planning. The other half of the class received teacher-led grammar and content advice for their 10-minute planning time. Planning included notes that were not to be used in the activity.

The planning time variables were measured for difficulty, which were three interactive tasks ranging in difficulty: (1) a decision making task, the most difficult task with a lot of new information; (2) a narrative task, the construction of a storyline from a set of five pictures; and finally (3), the easiest task, a personal information exchange task, involving directions. These tasks were done in pairs, with a listener and speaker.

The independent variables of planning time and task difficulty were then measured for the dependent variables of fluency, accuracy, and complexity. The researchers predicted planning would lead to fewer instances of dysfluencies, which they operationalised as reformulations, replacements, false starts, repetitions, hesitations, pauses and total silence; plus, higher complexity operationalized by syntactic variety of verb forms and the number of clauses per-c-unit. A c-unit can be described as an "independent utterance providing pragmatic meaning" (p. 310). That means a phrase, not necessarily

accompanied by a verb that carries a communicative value, for instance, in response to "Where is my jacket?", the elliptical answer "on the table", would count as a c-unit. Lastly, accuracy, or number of error-free clauses, was predicted to improve.

Fluency results in terms of the number of pauses and silence significantly decreased for the planning groups on all task types, especially for the narrative. Fluency was, as predicted, also lower for the difficult task, with that task producing more silence and pauses than the other tasks. Planning also led to increased complexity in the difficult task and more elaborate non-simple past tense use was reported in personal and narrative tasks, probably as a product of task type. Accuracy came out as the least significant variable in planning or task type, but the unguided planning group recorded most accuracy in all task types.

Overall more planning increased fluency and complexity, but accuracy results were not significant. Furthermore, although planning may help to negotiate difficult tasks, task type is a more complicated variable than at first appears.

Critical analysis

This paper has shown me that teacher-guided planning can help fluency and complexity, but that accuracy is better if it is unplanned and that task types can again produce different outcomes. Firstly, the results for fluency are encouraging, but when we re-check the definition of fluency we find the following: a "capacity to cope with real-time communication" and a propensity to use more "idiom-based language" (p. 304) because fluency and idiomatic expressions primarily drive meaning. This is a fair comment, but the extensive set of dysfluencies recording in this study do not take into account any aspects of idiomatic use for the measurements of planning, so the definition is inadequate. The dysfluencies also seem to be common features of L1 language production, and all may not be applicable to L2 speakers.

Next we turn to accuracy, often cited as a flaw of task-based teaching because of the importance tasks place on meaning itself. The authors use global measures, identified by the number of error-free clauses. This raises a problem of validity at the proficiency level of this study, pre-intermediate level, because similar errors would occur in high volume, and the study can not pick up more than one mistake in each clause, or what type of mistakes appear to be the most common, i.e. syntax, morphology or word order.

Both the variables of fluency and accuracy and, also, the methodology of using tasks for this study need to be examined. The three tasks used in the study represent only one end of a dichotomy starting at transactional tasks (Richards, 1990), or tasks that simply convey message or require completion. These tasks ignore the interactional properties of tasks to "create social interactions and social relationships" (p. 79). That is not so say that all tasks should be socially oriented, but some effort could be made to mirror what happens in the real world in order to benefit the language user. By ignoring the relationship side of interaction could lead to engagement in only artificial tasks that might suggest this study, and others, are merely to produce language for research purposes.

Measuring outcomes in pairs through tasks also presented a problem in this study. How much effect on dependent variables was due to joint planning or perhaps the work of one dominant individual, who may have instigated all of the conversation or turns necessary to complete the task? Indeed, what was the role of the listener? Was it the passive information receiver or constant confirmer of information? Conversation is an intricate network of turn-taking and to measure the results of pairs may, again, affect the validity of the results to establish the effectiveness of planning on oral outcomes.

Finally, since we are not told how many raters there were in this study reliability might be awkward, e.g. were repetitions made by a students due to self-monitoring or did the student use a repetition for rhetorical effect? It

would likely be the first situation, but use of a repetition for the latter would have been recorded as an error. Cross reference from more than one rater would have clearly solved such a problem.

In summation, this research addresses the complexity and diversity of tasks, and states that planning positively affects complexity and fluency, but it does suffer from problems of methodology and measurements of the oral outcomes.

*Article Two - **Review***

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

Rationale

The focus of this paper moves away from the classroom and measures students' oral outcomes on a semi-interactional proficiency test with various task types with limited planning time. The test was used to try and elicit the best possible response from speakers.

Study overview

Originally 107 students from various ESL L1 backgrounds sat a five-part proficiency test called the Australian Assessment of Communicative English Skills. Planning time was used as an independent variable on the five tasks in the test, ranging in rated difficulty as follows: (1) a conversation summary, rated most difficult; (2) a picture description and (3) picture comparison; (4) general discussion questions and; (5), rated easiest, a telephone answering message task.

During the test half of the students were given one minute for thinking and planning time on the picture and telephone message task, but no planning time on the summary or the discussion questions. The other half of the students had one minute planning time for the summary and discussion questions, but

no planning time for the picture or telephone tasks. The top 26 and bottom 28 students were omitted from both groups to focus on scores around a certain vocational/ functional level. This level was indicated in the paper as FSI 2 and 3, (but the author does not include any explanation of what this level means). The remaining 28 students from each group were assigned to a high or low group on the basis of their scores, creating a proficiency level independent variable.

The independent variables of task type, planning time and proficiency were measured with the dependent variables of complexity, fluency and accuracy. Complexity was operationalised by the number of clauses per text and number of subordinate clauses per t-unit(1) and researchers believed that complexity would increase with planning time.

Next, fluency was quantified by the number of self-repairs and repetitions per clause, and type token ratio. This ratio is the number of different words over a given segment. It was predicted that all fluency measures would increase with planning time.

Last, accuracy was measured by target-like use of verb morphology, including, verb agreement, tense marking, suffixes and plural morpheme - s, the last one, the researchers claimed, would not be affected by planning time. Also, included in the accuracy measures were the number of definite and indefinite articles. The author hypothesised that because definite articles are usually acquired before indefinite articles, the students in the no-planning group would revert to previously learnt forms, and that unplanned output would have more definite articles. Whereas the planning group had more time, and more opportunity to access later acquired indefinite articles during the test.

---

(1) A t-unit is one main clause, or other subordinate clauses that may be embedded in it (Crookes, 1990, p. 184). For example "I like chips, but not with any ketchup", would be 2 t-units.

I shall address the results, which were measured by two independent raters in the study, in terms of complexity, fluency, and accuracy for the lower and higher proficiency groups. The results for complexity show that planning enabled production of more subordinate clauses for the picture task and the most difficult task, the summary, but overall did not facilitate use of subordination in the easier tasks. The high proficiency group saw a greater amount of subordination than the lower group as a result of planning on the picture task.

As for fluency, the low proficiency group produced less self-repairs on the most difficult task, but did not provide added fluency in other tasks to any large extent. The high proficiency group monitored less with planning than the lower group on all tasks except the discussion questions. The other fluency measure, type-token ratio, only significantly increased for the low proficiency planning group in the conversation.

On all tasks planning helped accuracy to some extent with increased instances of target-like forms of verb morphology and decrease of non target-like use, but not significantly. Planning enabled only the high proficiency group to produce significantly more target-like verb usage for the most difficult task. Planning gains were also seen for high and low groups for the most difficult task and the high group in the next most difficult task. The lower group seemed to use more indefinite articles in easier tasks.

In summation, it appears that planning reduces some of the cognitive load for higher proficiency learners in more difficult tasks, but that lower proficiency groups may not be using planning time effectively. Accuracy and complexity seem to work together at higher proficiency levels, but that could be due to a perceived need to focus on form in the test situation. The results seem to further suggest that fluency may also benefit by planning.

Critical analysis

Two points that we need to ask about this study are, firstly, does limited planning time help students and, secondly, is analysis of test data helpful for answering my question from the introduction? Limited planning time was shown to help somewhat, but as seen in the previous study, this study, too, has failed to state how students used their planning time. To further investigate limited planning time, we also need to look at the dependent variables themselves, and at this point we see further discrepancies. In this study no explanation of accuracy, fluency or complexity is offered, so how can the author attempt to successfully operationalise, or even measure the output if it is not clear how the study approaches these notions? That leads into the problem of the measurements of the outcomes themselves. Fluency used only two measurements, self-repairs and repetitions, but in the previous study pauses, silence and hesitations were included. This study had no real way of measuring these three extra variables. How could the author differentiate between two forms of outcome that both contain the same number of these variables, but with differing levels of pauses and silence? As mentioned earlier, the author didn't provide enough information of how the variables were operationalised.

Following on from fluency, accuracy in this study favoured verb usage and definite articles. Although positive results were measured, nothing significant was found and that may lay in the choice of measurement. One measurement choice, indefinite over definite articles, seems a rather inefficient way of measuring accuracy, and no full justification for their inclusion can be found in the article. Without full justification and a small sample size (n=56), it is plausible that any arbitrary choice of grammar items, for instance possessive pronouns being more correctly used then personal pronouns, could have produced encouraging results in this study.

As for the test data, we need to look at the purpose of norm-referenced

proficiency tests like the one used in the study. These are designed to spread students out on a continuum of proficiency if students have little or no idea of what will be on the test. They should know the question types, but not what specific content or skills will be tested by those questions (Brown, 1996, p. 2). Norm-referenced have a specific purpose and to not recognise the assumptions of these tests is to invalidate them. Furthermore, these speaking tests should produce spontaneous language to mirror similar situations in real life as much as possible. In the last section of this test, students are requested to give answers to discussion questions. Does planning on this test mean that these test takers knew which questions were coming or what general topics would be discussed? Both situations would have an impact on results.

In summary, then, the fluency, accuracy and complexity recorded some gains due to planning, but the measurements in a testing situation may be unreliable and seem to be less valid for answering my question. As regards task type and proficiency, we can conclude that higher proficiency learners benefited most from planning, especially for more difficult tasks. In explanation of the lower proficiency group's apparent lack of benefit from planning time, even the author simply admits that planning time should be considered in some activities, but not others. I would go along with this point and add that perhaps the planning time of 1 minute was too short to make any real difference at all.

Article Three - Review

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. ***Studies in Second Language Acquisition, 20, 83-109.***

Rationale

This study moves back to the classroom and appears to try and find some

middle ground between the last two studies for planning time by providing differing planning time: ranging from 1 to 10 minutes to concentrate on two separate activities.

Study overview

The subjects were 31 intermediate level German language learners from various L1 countries in a foreign language environment at a London university. These students were randomly assigned to one of four groups, and each student from the groups individually performed two taped tasks over a two-week period. Task one was described as a personal task and involved using personal information, in this instance leaving directions on a telephone answering machine. The second task, a narrative task, was more difficult than the first and was concerned with constructing a story in the past tense from five specified words (the types of words were not mentioned, i.e. nouns, verbs, etc.). Planning time was used as an independent variable in the four groups. Those groups were separated into one control group that had no planning time and the other three experimental groups had varying degrees of planning time at 1, 5 and 10 minutes respectively, in which time for taking notes in German was allowed. Students, however, could not use these notes to read from during the task proper. Along with planning time, task type was also measured for the dependent variables of fluency, complexity, accuracy and lexical density.

Fluency measurements were number of pauses, total pausing time, mean length of run (mean number of syllables between pauses) and two speech rates. Complexity was measured by words per c-unit, number of subordinate clauses and s-nodes per t-unit. Unfortunately, the author offers little in the way of a description of what an s-node is. Finally, accuracy was determined by the number of errors per 100 words, number of error-free clauses, and a count of word order and lexical errors. In addition to fluency, complexity and accuracy,

lexical density, a new dependent variable was used in the study. Lexical density was a calculation based on the number of high and low frequency words used. In the calculation less weighting was given to high frequency words that appeared more than once in a speech sample. This weighted number of high and low items was presented as a percentage of the total number of words. The researcher hypothesized that all four dependent variables would improve and the more difficult task would become easier with more planning time.

Planning improved fluency and lexical density especially for the personal task, with the main differences seen between the no planning and 10-minute planning group. On the narrative task, planning time increased fluency, but no major differences were viewed among planning time groups.

Both tasks resulted in more accuracy produced by the planning groups than no planners, especially in number of errors. Words and lexical choice error improved slightly with planning. The narrative task saw gains from the no planning group to the 1-minute group, but the number of errors, surprisingly, increased with planning time. Accuracy seemed to become more important to students at the 1-minute planning stage, whereas complexity benefited from 10 minutes planning. That is, accuracy and complexity are competing for the attention of limited resources at different times during output.

On the whole, planning produced more fluent, more lower-frequency lexical items and higher complexity the more planning time was available, but accuracy seemed to be more important for planners with less time.

Critical analysis

Longer planning time seems to help fluency and complexity more than accuracy, but as this study only used two particular tasks generalization to all spoken activities would be difficult. This study did, however, suffer slightly from problems found in previous studies, and problems of validity and

reliability.

As with other studies, a definition of the outcomes of fluency, accuracy, complexity and a fourth outcome, lexical density, is necessary and it would have also been very useful to know the mental state of the learners through a qualitative study.

The author claims that roughly 30% of German words undergo inflection, which can be based on case or gender. With this in mind a deep understanding of the difference between German and English languages is necessary. However, the author chose not to include such a thorough explanation of these differences, signalling external validity problems. Explanations are also deficient in one of the measurements used for complexity.

Finally, the reliability of this study may have been undermined by the variation of the groups involved. The only way that the author could group these students was by a previous proficiency grade, at least one semester old. There is no way of telling what the students had done in the way of study or been exposed to in the interim that may have affected their performance and the subsequent reliability of the test.

Generally, the study seems to conform to the theories mentioned in the introduction of this article about competition for available resources at various stages of planning and does help me answer my question about which outcomes are better as a result of differing planning times. Having said that, a variety of different activities and an insight into mental processes, plus more description of the accuracy measurements, could have told me more about how the goals of complexity and accuracy seem to work against each other.

*Article Four - **Review***

Ortega, L. (1999). Planning and focus on form in L2 oral performance. ***Studies in Second Language Acquisition, 21, 108-148.***

Rationale

This study focuses on one task type to measure four dependent variables of accuracy, fluency, complexity, and as in the last study, lexical density. Planning time was at zero and 10 minutes, but unlike other studies, the author has attempted to find psychological explanations for the oral outcomes by using questionnaires and interviews. Listener comprehension has also been built in.

Study overview

The subjects were 32 pairs of American students of varying ages, who were advanced Spanish L2 speakers in Hawaii. These students were randomly assigned to four groups with planning time used as an independent variable. Two control groups received no planning time and the other two groups had 10 minutes planning time. Two narratives similar in length and complexity were first played on a tape to the control and experimental groups, who had access to a storyboard of eight pictures. The control groups were required to retell the story to a listener immediately and the experimental groups had 10 minutes planning time in which they could make notes that were not to be used in the activity before they had to explain their stories to a partner. The partner was required to complete written tasks during and after the activity to show that he or she understood correctly what was said.

The independent variable of planning time was quantitively measured for complexity, lexical range, accuracy and fluency. Complexity was measured by words per utterance: a stream of speech under one intonation contour that is bound by pauses (p. 124). Lexical range was calculated by type token ratio, the number of different words divided by the total number of words. Accuracy was measured by target-like usage of noun modifiers and articles, and fluency was analysed by syllables per second arrived at by removing self-corrected words, repetitions and pauses over 3 seconds and divided by the total articulation time. After each activity all the subjects took part in interviews to

try to establish individual psychological processes of planning.

The quantitative results showed planning helped all dependent variable measures. However, significance was only found in fluency and complexity. The increase in type token ratio for lexical range of the planning group was not significant. Lastly, accuracy had mixed results, with target-like usage being significantly higher for the planned group, but increases through planning in target-like use of articles was not found to be significant and the author concedes that results for accuracy may not be immediately noticeable.

The qualitative data revealed that the speakers used a variety of problem solving strategies, with over half interviewed using strategies like rehearsal and talking or reading to oneself. Memory of grammar and vocabulary was also better facilitated by having written notes during planning. Other speakers claimed that because the task itself required them to communicate a message, the needs of the listener figured prominently in the organization, possibly incorporating a need to lower the difficulty of lexical and grammar items to ensure that comprehension took place.

Planning also allowed students to monitor, or correct mistakes, during the planning and output stage, either by auditory or visual means. All of the above approaches by students show a complicated relationship between planning and focus on form. On the other hand, some students saw the planning as unnecessary because of the ease of task and the fact that notes couldn't be used in the task. Lastly, the no planning group seemed unhappy with their fluency and felt that immediate output made them speak stiltedly.

In summary, the quantitative results indicate that planning helped fluency and complexity, but not lexical range, and accuracy is once again a slightly unknown element. The qualitative results seemed to suggest that some people focus on forms, but other factors such as individual differences, proficiency, and task complexity may moderate the effect of planning.

Critical analysis

This study offers both a first step into telling us how the mind of the speaker works when initiating planning, and also, encouraging results for complexity and fluency. However, there are problems with reliability of the methodology, oral outcome measurements, and the qualitative results themselves.

Firstly, methodology problems are related to the practice effect that students would have experienced by taking the tests so closely together. Although the author has used counterbalancing, the similar nature of the tasks could have caused an improvement the second time around. Also, the high proficiency-level volunteers were grouped on the basis of information from a test at least one semester old. Who knows what language or experience learners of this proficiency might have been exposed to in the interim? Also, in the methodology the students were told to not write full sentences during planning time, but no reason for why this point was made has been offered. If the researcher simply wanted to find out how students approached the activities, why did she try to influence part of the outcome by issuing an instruction to not write in full sentences, as it may have altered the level of accuracy achieved in the results? Having criticised the methodology for its unreliability, it is worth mentioning the benefit of incorporating measurement of the listener's comprehension of what the speaker said. This added element has introduced a reason for speaking and proof that comprehension had taken place. The last point has been lacking from the previous three studies.

In this study problems of oral output measurement can be traced to accuracy measurements. Information is lacking about the decision as to why the author chose to measure accuracy by analyzing noun modifiers and articles. These seem to be chosen arbitrarily, and a much more elaborate explanation of the choice of these measurements and their relation to accuracy should have underpinned this study.

The qualitative information that the author offers does give good insight into possible strategies that teachers could teach to less proficient students at the planning stage to help them negotiate speaking activities. Answers taken from students about how they approached the activities, such as not taking unnecessary risks or sorting out essential from the inessential, would help guide some students into producing better quality output. However, presentation of the interview information could have been more thorough. We are told, for example that some speakers considered accuracy more important than communication, but not how many of the speakers set about the task with the idea in mind. This could be central to the issue of which resources the speakers focus on when they engage in speaking activities.

Summing up, planning does help fluency and complexity and this article can help teachers use planning time more effectively, but information contained in this report needs to provide more in depth explanation and some of the results need to be questioned because of the lack of care taken to protect reliability.

*Article Five - **Review***

Yuan, F. & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. ***Applied Linguistics 24 (1), 1-27.***

Rationale

This study concentrated on only a narrative task type for three reasons. Firstly, to ensure monologic speech without partner interference, secondly, to produce a reasonable difficulty level and, thirdly, to make comparison easier with other studies. This study introduced a new independent variable called on-line planning. As an alternative to planning before the task (pre-planning), on-line planning implies a focus on form while engaged in the activity itself for an unspecified length of time. On-line planning was perceived to relate the

formulation (emphasis on grammar) stage of the stage of Levelt's models cited in the introduction of my paper.

Study overview

This article studied 42 Chinese university English majors aged between 18 and 20, with similar proficiencies based on their TOEFL scores. These students were randomly assigned to three different independent variable planning groups. The control group had 30 seconds planning time before being asked to finish the task within 5 minutes. The next group were given 10 minutes pre-task planning time to make notes only, not to be used in the task and 5 minutes for actual task completion, so the no-planning and 10-minute group were both under time pressure. The third group was the on-line planning group with 30 seconds to prepare and unlimited output time. All of the students in these groups had to individually record a narrative requiring four sentences for each of six pictures on a storyboard on to a tape.

The planning times were then measured for fluency, complexity and accuracy. Firstly, fluency was measured by two rates, one was syllables per minute, a measure of the number of syllables in a story divided by the number of seconds taken to complete the task. The other rate was the same as the first, but all words or phrases repeated, reformulated or replaced were removed. Fluency was expected to increase more with the planned group and to impede the on-line planning group.

The next variable, complexity was calculated by number of clauses per t-unit for syntactic complexity, type-token ratio for lexical variety and syntactic variety, a calculation of the number of different grammatical forms, including tense, modality and voice. Complexity was expected to increase with both planning and on-line planning.

The final dependent variable, accuracy, was measured in two ways: (1) the number of error- free clauses, and (2), a percentage of correct verbs regarding

tense, aspect, modality and subject agreement. The researchers speculated that planning would lead to little change in accuracy, but that accuracy would increase as a result of on-line planning. Again, as in the last study, to complement the quantitative data, subjects were examined qualitatively via a questionnaire and interview to try to establish how they planned and organized their stories, and how they approached grammar and vocabulary during both planning and the activity.

The results of this study have been summarized in Table 1, with one

Table 1: Planning results (from Yuan & Ellis, 2003)

Task condition	Fluency	Accuracy	Complexity	Lexical
No planning	2	3	3	2
Pre-task planning	1	3	1	1
On-line planning	3	1	1	3

indicating the highest performance out of the planning groups and three, the lowest, for each dependent variable. In terms of overall time, the on-line groups spent more time on task than the no-planning and pre-task planning groups, and that the on-line and pre-task groups produced more language. The pre-planning group was more fluent for both rates, and the no-planning group was slightly less fluent than the on-line group. This last group spent a lot of time reformulating and repeating phrases. T-unit scores for complexity were similar for on-line and pre-planning groups, both of these groups produced speech that was significantly more complex than the no-planning group. Syntactic variety provided similar pre-task and on-line planning scores, but those differences were not significantly different from the no-planning group. Next, lexical variety showed the most difference between the lower scoring on-line planners and the higher scoring pre-task planning group. On-line

planners came out on top for the two accuracy measures, which were both significantly higher than for the no planning group.

Some of the qualitative interviews seemed to add weight to the above results with planning-time students concentrating on creating outlines of what they wanted to say, possibly leading to production of more lexical variety and fluency to the detriment of accuracy, conceptualization over articulation, or a need for lexical meaning over syntactic form. The on-line group were stronger in accuracy measures because they seemed to favour simpler vocabulary in order to concentrate on syntax that extra planning time during the activity provided. This extra time was used to access syntactic constructions from the long-term memory, again all pointing towards students prioritizing what was needed for the completion of a task.

Critical analysis

This study seems to have succeeded in creating better conditions for measuring the dependent variables by removing partner interference, as well as tentatively confirming all the hypotheses. However, this article does seem to be narrow in scope and there are other problems with the accuracy measure, the practicability of the study, and the questionnaires and interviews after the activity.

To begin with, there were only three groups and those group members only worked on one particular type of task, which was telling a story. Initially this was done to draw comparison with other tasks, but it would have given this study extra credibility to have had more varieties of spoken activities, so we could see the range of on-line planning outcomes or at least two groups doing each type of planning leading to comparison and reliability checks.

Next, the global accuracy measure, as in previous studies, left no way of accounting for clauses with more than one error. Indeed, this study defined errors in clauses as phonological, syntactic or lexical, but nothing about type of

individual clause errors was mentioned.

Finally, this study utilized questionnaires and interviews to look at the psychological mechanism of planning. However, information from the questionnaire is mentioned only once and that pertains to how the majority of students made an outline during planning time. No additional information is given for the processes underlying either the no-planning or on-line planning groups, the last of which was the main focus of this study. We are left wondering if the points raised in the discussion part of this paper are conjecture, or based on concrete examples from student feedback.

Generally, this study deals adequately with creating the right conditions to properly measure oral outcomes, but a narrow activity focus and lack of pertinent information detracts from the informative analysis in this paper.

Conclusion

In answer to my question in the introduction, it is clear that planning is effective in improving the quality of spoken output of language learners. However, individual differences between students and their approaches to doing things like spoken tasks can lead to priority being put on either a focus on meaning or a focus on the grammar element, focus on forms. Speaker planning must address the composite parts of focus on form, namely: accuracy, fluency, complexity and lexical density. In addition, these parts need to be produced at equal levels.

The studies featured in this paper are the SLA field's attempt to try and explore the notion of planning at a deeper level and exactly how it affects output, what manner of priorities of focus on form are taking place, and what other factors can also influence this process.

When describing output, it is important to establish exactly what we mean when we talk about accuracy, fluency, complexity and lexical density. Only two of the articles featured in this paper have tried to answer this question.

Without a full description of exactly what each of these elements involves, studies cannot hope to successfully operationalize them for measurement. Beyond the four dependent variables of focus on form used in these studies, Celce-Murcia, Dornyei & Thurrell (1997) call for an expansion of the notion of focus on form into the areas of discourse rules, pragmatic awareness, strategic competence, and lexical formulaic phrases (p. 147). That is, there may be a lot more than four elements to consider when measuring focus on form.

Operationalization leads into measurement, which must also be precise. Some of the measurements used in these studies led to inconsistencies and reliability problems. Although I believe that a variety of measures have been found for fluency and complexity that seem to adequately describe what the students are producing, accuracy has suffered slightly. The studies have moved from being general, i.e. a count of error-free clauses, to being too specific, for example measuring -s plural morpheme or articles. What seems to be needed is a compromise of global measures that can then be grouped into error types. These error types could then be weighted using partial credit modelling (Bond & Fox, 2001) to create a more clear-cut picture of accuracy.

The last two studies in this paper have also tried to tap into the mind of the speaker and seek an explanation for focus on form decisions by canvassing the students after completing the tasks. This is essential to understanding how students prioritise form and meaning, but results were not thorough enough, with little mention in the studies or speculation about underlying psychological processes that lay behind decisions of prioritising on one aspect of language over another. This kind of study could be fertile ground for the future as long as researchers can correctly identify the components of focus on form.

Next, we can look at the results themselves. Overall, fluency has seen the best improvements as a result of planning. Indeed, fluency seems to increase with amount of time given for planning. Complexity also has received some promising results that seem to favour planning over a longer period of time,

especially with higher proficiency learners. Could it be that the complexity and fluency constructs are similar in nature because researchers have predicted and found a linear relationship between complexity and fluency as a result of more planning time. For lexical range, results have been mixed with one study showing positive significant improvement with planning time and another study recording only minor changes. Lastly, what then is happening with accuracy and planning time? Results have been very mixed, but there seems to be a trend of results that suggests that the goal of accuracy becomes prominent with less planning time, and as the planning time increases students start to think more about complexity. Also, higher proficiency learners can attend to accuracy better than lower proficiency learners for more difficult activities. To further strengthen the reliability and provide extra insight of these results, researchers need to provide full and precise description of the measurements involved.

Along with planning and outcome variables these studies have incorporated many other variables. These have included task complexity, task type, learner proficiency, length of planning time, a testing situation and interaction through task-based medium. Some other possible variables not mentioned here, but those that might require exploration could be related to output in different oral situations, like presentations or speeches, both of which use speech that is not influenced by other speakers.

At this point I would like to turn my attention to a number of the variables mentioned above that perhaps have more of an impact on my question in the introduction. The first is the use of tasks when measuring the effects of planning. There is much research to date regarding the positive effects of tasks and how they can bring about co-operative group learning and help with acquisition (Long & Porter, 1985), as well as being so diverse in nature that any teacher can incorporate them into any type of class for different purposes. However, it is because of the interactive nature of tasks that I believe them

to be inappropriate for measurement of oral outcomes. It is difficult to build reliability into a study if one particular partner is dominating the conversation and producing all the turns. Studies that have excluded the interaction element from the speaking situation and purely left the speech of individuals, have, to my mind, succeeded in at least partly accurate measurements. If tasks are to be used for analysis one possible study could involve measurement of individual outcomes after exposure to things like cognitive and metacognitive strategies, task-based curriculum, or direct teaching (Richards, 1990 p. 79-81) approaches that teach explicit conversation skills.

The testing situation in the second study also deserves attention. Testing and normal classroom situations are different because both have different objectives. Giving planning time in testing situations should not allow students access to information contained in the test and therefore risk reliability issues. To answer my question, I believe it is important to keep the purposes of testing and general measuring of outcomes separate, so as to not confuse one with the other.

In conclusion, research into planning as an essential element of classrooms is in the beginning stages. There are many factors outside of just planning itself that can influence oral outcomes, so efforts need to be made to define exactly what needs to be measured in the outcome, how it is to be done effectively, and how to find the optimum situations for measurement.

References

Bond, T. G., & Fox, C. M. (2001). *Applying the rasch model - fundamental measurement in human sciences*. Mahwah NJ: Lawrence Erlbaum Assoc.

Brown, J. D. (1996). *Testing in language programs*. New Jersey: Prentice Hall Inc.

Celce-Murcia, M., Dornyei, Z. & Thurrell, S. (1997). Direct approaches in L2 Instruction: A turning point in communicative language teaching? *TESOL Quarterly,* [*Vol.*] *31* [*1*], 141-152.

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics, 11 [2],* 181-197.

Gass, S. (1997). *Input, interaction and the second language learner.* Mahwah, NJ: Lawrence Erlbaum Associates.

Long, M. H. (1991). Focus on form: a design feature in language teaching methodology. In de K. Bot, D. Coste, R. Ginsberg, and C. Kramsch (Ed.). *Foreign language research in cross-cultural perspectives,* (pp. 39-52). Amsterdam: John Benjamins.

Long, M. & Porter, P. (1985). Groupwork, interlanguage talk, and second language acquisition. *TESOL Quarterly 19* [2], 207-228.

Long, (1996). The role of linguistic environment in second language acquisition. In W. Ritchie and T. Bhatia (Eds.). *Handbook of second language acquisition,* (pp. 413-468). CA: Academic Press.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11,* 113-128.

Richards, J. (1990). *The language teaching matrix.* Cambridge: Cambridge University Press.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 17-46.

Van Patten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12,* 287-301.