

Effects of Planning on Language Learners Writing Performance: A Case Study

Andrew PHILPOTT and Abdullah ALAMI

It is a commonly accepted notion that planning may improve a second language learner's writing performance since the human capacity to process whilst also maintaining accuracy is quite limited. However, the effects of pre-planning as opposed to no planning have not been documented very well and are generally of a vague nature. It is therefore important to be able to observe and accurately measure the effects of extra planning time on a second language learners writing performance. This small-scale study, implementing the test-retest design, assesses the effects of pre-planning time on the writing performance of two second language learners using both qualitative and quantitative analysis. The primary data source used was two diagnostic essays produced on two separate occasions. Factors that were considered as part of the analysis process included qualitative ones such as overall essay structure and the cohesiveness of sentences and ideas whilst the quantitative analysis focused mainly on measurements of Complexity, Accuracy and Fluency (CAF Measurements).

INTRODUCTION

The study of the effects of planning on second language learners' writing performance is of interest to researchers and teachers. Testing the theoretical claims of the effects of different models or variables should be conducted via empirical means. One such claim is that allowing participants more time for planning before a writing task would help them better focus on form, which would also have pedagogical implications (Doughty & Williams, 1998). Schmidt and Frota (1986) argued that this extra time for planning would allow the participants to notice the gaps in their repertoire. Swain (1998) suggested similarly that it would allow students to notice holes in their inter-language with the hope that ultimately they could learn from it. The study of the effects of no-planning vs. pre-planning can help provide information as to how a simple task implementation condition, such as the allowance for pre-planning time, can be manipulated to facilitate a more balanced development in terms of linguistic complexity, accuracy, and fluency.

Humans, unlike the super computers of today, are limited in their processing capacity; therefore the amount of conscious attention available for the processing of incoming data is limited (Crookes, 1989). The portions of

processing capacity working on complexity or fluency might not be able to be used to concentrate on accuracy at the same time. As such, conscious attention to both information content and form simultaneously is quite difficult; and as a result most people have to decide for themselves where they prioritise their attention (Skehan, 1996). This study focuses on manipulating the task implementation condition by simply providing for an extended pre-planning time in one task whereas none is provided in the other task. This enables researchers to explore the impact of pre-planning time using both qualitative and quantitative means in the context of producing a written essay.

METHODOLOGY

The study used a test-retest design, in which the participants were tested twice, to determine the effect of the selected variable. The task that the participants were required to perform was simply that of writing an essay. This essay writing task served as the diagnostic test. The main aim of this writing task was to collect data for this report and to illustrate the effects of the variable (25 minutes of pre-planning time). Though the essay question between tests was slightly altered, the overall essay structure, as well as the experimental design and measurement, remained the same.

Participants

The data of two participants was used in this case study. The two participants were selected from a pool of four participants after the completion of a fifteen-minute interview. The interview was carried out to find out information in relation to their backgrounds, their ages, their IELTS scores, their current areas of studies, and their current English abilities. The two specific participants selected for this study both had the same IELTS score of 7, which indicated they had a good command of the English language. Furthermore their PhD-level studies at the University of Melbourne demonstrated their ability to cope with the rigours of academic life. The students were not noted as being at risk by the university nor did they themselves express any particular area of English usage that they wanted formal instruction/tuition in. The other two participants were deemed inappropriate as they would have introduced additional variables to account for, due to their different backgrounds, different cultures, different genders, varying IELTS scores and current academic levels. One of the two also stated they had had issues whilst studying and required formal support. The two selected participants were both male and in their mid-30's, one from Japan and one from Korea. Both had spent time in each other's countries and understood the languages and cultures of both countries. They were both linguistics PhD candidates at the University of Melbourne.

Data

The primary data sources were the essays produced by the participants on two separate occasions. The two essays, functioning as diagnostic tests, were

kept comparatively simple, similar in topic, and open in scope to elicit the greatest possible response. Both essays were completed using pen and paper. The first essay topic was given to both participants at the same time with no prior knowledge of the expected content or structure. It was a timed task and whatever was produced was collected at the end of the 20-minute time period allowed for its completion. The participants were free to use the 20 minutes as they deemed suitable although great emphasis was placed on writing as much as possible within the allocated time. The second essay had a 25-minute pre-planning time allowance in addition to the 20-minute writing limit but was again written with no knowledge of the assigned essay topic beforehand.

In both essays, the participants were observed and supervised in exam-like conditions with no interaction or collaboration between the two. In order to reduce the many variables, the location, the conditions, the participants and the essay questions were carefully considered. It was vital to focus on the writing and linguistic ability of the participants. Therefore, simple topics were chosen that they could both write about, keeping the content focused. Both topics were designed to encourage the participants to produce descriptive or argumentative essays. The first essay involved writing a response to the prompt *Write about your favourite city* whilst the second essay required a written response to the prompt *Write about Melbourne*.

On completion of the first (no pre-planning) essay, participant feedback was sought via a mini-interview in which both participants expressed the desire for more time in general, and specifically planning time. The second (pre-planning) essay, which had a 25-minute planning section, was designed to allow for the formulation of their strategies and focusing of their thoughts prior to the 20-minute writing task. On completion of this second essay their feedback was again sought to help with the qualitative analysis. The participants did not have any prior knowledge regarding the content or structure of any of the essays which meant they had no idea of what to expect. Furthermore, no hints or help of any sort was offered to any of the participants upon commencement of the essay task.

Data analysis

The essays produced by the participants in the no-planning and pre-planning tasks were analysed using a variety of quantitative and qualitative methods. As part of the analysis and examination of the written essays, the salient features were also considered. Two of the many reasons these two particular participants were suitable for this study, were the similarity in their IELTS scores and their educational backgrounds (currently PhD level students in the same field). Though this does not necessarily imply that they have identical writing proficiencies in English, it does mean they are similar in profile.

Qualitative analysis

Qualitative analysis is not indicative of a specific research stage but rather begins from the very inception of research including the choice of participants, observations and activities involved (Richards, 2003). As there were only two participants, the qualitative analysis consisted of examining the overall essay structure (e.g. introduction, body and conclusion), the cohesiveness of the sentences and ideas presented, and the effectiveness of the argument and its relation to the conclusion. Further qualitative data was gathered from the two interviews conducted during the study.

Quantitative analysis

The quantitative analysis of the produced texts was carried out using CAF measurements (Complexity, Accuracy, and Fluency). As written texts, such as these essays, contain multiple components, the principal components may possibly be captured by the notions of CAF. In recent years the CAF has emerged as a notable complement to other established proficiency models such as the traditional four-skills model and the sociolinguistic and cognitive models of L2 proficiency (e.g. Bachman, 1990).

It is important not to confuse the complexity being referred to in this report with that of cognitive complexity. Although the term complexity in second language acquisition literature often implies both cognitive and linguistic complexity, they are not the same. This study focuses on linguistic complexity. The complexity measures used in this report allow for differentiated measurement of linguistic complexity by using both grammatical complexity measures and lexical complexity measures. In this paper grammatical complexity was measured using the clauses per T-unit(C/T) and dependent clauses per clause (DC/C), methods. The combination of these two methods should give a good overall picture of the grammatical complexity employed, including subordinations and embedments. Lexical complexity can present an indication of the complexity of a written piece and is an important aspect of academic writing. However, there are a variety of methods by which complexity can be measured, so it may be necessary to use a variety of measures as complexity is complex (Norris and Ortega, 2009). Lexical complexity was measured using the type-token ratio and the overall percentage of words found within the combined Academic Word List(AWL)and General Service List (GSL) corpuses.

Accuracy can be measured in various ways (Polio, 1997; Wolf-Quintero, Inagaki, & Kim, 1998), although Bley-Vroman(1983) argues that accuracy is not even an indicator of inter-language development. In this report accuracy was simply a measure of the total number of errors per total number of words (E/W) (providing a holistic overview as opposed to ratio scores which cannot differentiate between T-Units containing multiple errors as opposed to a single error) and the number of error free T-units per T-unit (EFT/T). Wolfe-Quintero et al. (1998) acknowledge the disputable transient nature of accuracy

measurements since “the purpose of accuracy measures is precisely the comparison with target-like use. Whether that comparison reveals or obscures something about language development is another question” (p. 33).

Fluency was measured using Microsoft Word to determine the total number of words (W), and the words per T-units ratio (W/T). A T-unit was defined as a main clause plus any number of dependent clauses embedded or attached (Hunt, 1996).

In order to reliably analyse the texts using these measures, it was necessary to analyse the text coding for clauses, errors and T-units. A clause can have a variety of definitions and what one researcher may consider a clause may differ from another and these actions thereby have their own research implications (Wolf-Quintero et al, 1998). However, for the purpose of this paper a clause includes either the dependent or independent clause, including infinitive clauses.

Errors were classed using Chandler’s (2003) guidelines and all minor errors such as spelling mistakes, punctuation errors and missing/unreadable characters were ignored. However, incorrect uses of articles, singular/plural nouns, incorrect joiners (or lack of) as well as verbs were counted. An incomplete sentence at the end of the text was not considered an error as this could have been a result of time having finished.

The Academic Word List (AWL) developed by Coxhead (2000) and the General Service List (GSL) developed by Michael West (1953) were consulted to cross reference words for frequency as a measure of lexical complexity. The combined lists contain some 570 word families and about 2000 words from a corpus of written academic English.

TABLE 1
Quantitative measures used in assessing and analysing the essays (CAF)

Complexity	Accuracy	Fluency
<i>Grammatical Complexity</i>		
C/T	E/W	Words
DC/C	EFT/T	T-Units
		Words/T-Units
<i>Lexical Complexity</i>		
Type-Token Ratio		
% words in AWL and GSL		

FINDINGS

Qualitative analysis

The essays with no planning either started with a question or a brief statement. They contained numerous strikethroughs indicating the self-corrections that participants made. In comparison the essays with pre-planning started with a general introduction, outlined the viewpoint of the author and gave a sense of what the essays were going to discuss in a more logical manner.

The pre-planned text was written in a well-structured manner with an introduction, body and a conclusion. The structure did seem formulaic without much deviation. The participants in the post-test interview did indicate that that was the pattern they were taught and they did not want to venture far from it. Visually the handwriting in the pre-planned texts was more composed and orderly than those of the no-planning texts.

The excerpts below illustrate the differences between no planning compared to pre-planning. Here is a no-planning excerpt:

There are a lot of cities I would like to describe as a favourite city. I like to choose Seoul, South Korea as the most favourite city of mine.

As opposed to a pre-planned excerpt:

Melbourne is known as the most livable city in the world according to a survey conducted by a British newspaper "Economist". It has numerous elements that attract tourists from all over the world ...

The pre-planned essays were also much longer having increased from an average word count of 212 to 391, representing a 54% increase. The text produced also appeared more sophisticated in design and wording. The focus was also more consistent and the structure more cohesive with the use of more appropriate conjoiners. The direction and flow of the arguments were also clearer, better substantiated and easier to understand. The following excerpts, from the same participant, illustrate this improvement in sophistication, argumentation and substantiation.

...my favourite is Melbourne because I can have many opportunities to meet and interact with people from various countries and can feel exostic from the city.

As opposed to

Melbourne is a diverse and multicultural city in the world. According to Wikipedia, it was noted that 35.8 per cent of the population was born overseas, exceeding the national average of 23.1 in 2006. ..The cultural diversity is reflected in the city's restaurants serving various international cuisines not surprising considering over two-thirds of Melbournians speak only English at home (68.1%), followed by

Chinese, Greek, Italian and Vietnamese.

Between the two tasks certain observational differences were also noted. One observation was that one of the participants (the Korean participant) spent one-third of the time in the no-planning task contemplating as opposed to writing anything. Another observation was that the participants in the pre-planning task, as opposed to the no-planning task, appeared to be calm, collected and writing at a steady constant pace. In mini-interviews immediately after the first essay task, participants felt disadvantaged by the time constraint. One participant commented, "I can't go back and review what I wrote." These observations are all consistent with Polio, Fleck and Leder's (1998) study and Power and Fowell's (1996) study expressing the need for more time so that participants do not feel disadvantaged.

Quantitative analysis

TABLE 2
Results for Complexity

	Mean	SD	Min	Max
Grammatical				
C/T				
No Planning	2.00	0.29	1.79	2.20
Pre-Planning	2.36	0.24	2.19	2.53
DC/C				
No Planning	0.41	0.06	0.36	0.45
Pre-Planning	0.43	0.06	0.39	0.47
Lexical				
Type/Token Ratio				
No Planning	0.55	0.002	0.55	0.55
Pre-Planning	0.53	0.050	0.49	0.56
% words in AWL/GLS				
No Planning	87.66	5.01	84.11	91.20
Pre-Planning	86.67	5.10	83.06	90.27

Grammatical and lexical complexity

The grammatical complexity results of the pre-planning essays exhibit a slight improvement in comparison to the no-planning essays (refer to Table 2 below). The mean C/T count increased from 2 to 2.36 whilst the mean DC/C count increased by a mere 0.02 from 0.41 to 0.43. It is important to remember

the limitations of this study and why this might have occurred (as noted in limitations section). On the other hand the mean lexical complexity took a slight fall on both measures.

Accuracy

A summary of the findings is outlined in Table 3 below. The results indicate that where the errors-per-word-count (E/W) and the error-free T-units per T-unit count (EFT/T) is concerned the participants' accuracy increased when the participants had time to plan their essay. The Mean E/W measure reduced from 0.05 for no planning to 0.03 for pre-planning. This increase in accuracy is also corroborated by the EFT/T measurement. The mean EFT/T measure increased from 0.29 to 0.49 (note that one of the participants more than doubled their EFT/T). On both measurements this indicates an increase in accuracy.

Some of the errors that were present were due to incomplete sentences, typically consisting of only a subordinate clause or an incomplete idea. Ultimately, it is important to note that there may be differences that are simply not measurable accurately enough (Knoch & Elder, 2009). And, that the effects of planning can have mixed results on accuracy, as its influences are different from person to person (Ellis & Yuan, 2004).

TABLE 3
Results for Accuracy

	Mean	SD	Min	Max
E/W				
No Planning	0.05	0.01	0.04	0.06
Pre-Planning	0.03	0.01	0.02	0.03
EFT/T				
No Planning	0.29	0.01	0.29	0.30
Pre-Planning	0.49	0.19	0.35	0.62

Fluency

Fluency was measured using three indicators: the total number of words, total number of T-Units and total number of words per total number of T-Units as shown in Table 4. What is interesting here is unlike other studies such as Storch (2009), the mean total number of words had substantially

increased as had the T-Units whilst the Words/T-Units had increased at a lesser degree. The standard deviations for words were quite large on both counts which indicated there was a large difference between the students.

TABLE 4
Results for fluency

	Mean	SD	Min	Max
Words				
No Planning	212	24.04	195	229
Pre-Planning	391	94.75	324	458
T-Units				
No Planning	12	2.83	10	14
Pre-Planning	19	2.83	17	21
Words/T-Units				
No Planning	17.93	2.22	16.36	19.50
Pre-Planning	20.35	2.07	18.88	21.81

DISCUSSION AND CONCLUSION

It is vital to point out the limitations of this study. First and foremost, there were only 2 participants. Such a small sample size can in effect show almost any possible result, and therefore cannot be considered to be statically significant. So, a much larger sample size is required to be able to deduce more meaningful information. Due to this small sample size, no further statistical analysis would have been reasonable or more meaningful, such as hypothesis testing, t-tests, ANOVA, etc. Even calculating the values for the mean, standard deviation, minimum and maximum was arguably unwarranted but was completed to show some basic trends.

Secondly, though every effort was made to ensure the participants were as similar to each other as possible both in their language ability and in their

background, it should be obvious they were at differing linguistic levels; despite the similarities in their backgrounds, they were not the same.

Lastly, the third limitation is the quantitative measurement used. CAF also has its own limitations including the very definition of CAF as being a scientific measurement, the effects of learner variability's, the effects of pedagogical interventions, the effects of task complexity and even contextual factors such as the characteristics of the input itself (Housen & Kuiken, 2009).

The overall findings of this study are in keeping with those of Caudery (1990), Hale (1992), Kroll (1990), Livingston (1987), and Powers and Fowles (1996), all of whom found some differences in results, according to various measurements, for essay tasks produced under different time conditions and with pre-planning. Similar to their results, the longer pre-planning task did yield better results in a number of specific measures such as a slight increase in the mean scores for grammatical complexity (C/T and DC/C), an improvement in accuracy (EFT/T and E/W) (which is contrary to results in Storch (2009)), and fluency with increases in all three measurements used, though most significantly in word count. Qualitatively, there was evidence of greater planning, better cohesiveness, better structure and better presentation of ideas. From an observational standpoint the participants also appeared more calm and collected and wrote at a steadier pace in comparison to the no-planning task where the Korean participant spent one-third of the time contemplating the actual question. This further reinforces the benefit of pre-task planning time in line with Van Patten's (1990) theory that pre-task planning opportunities allow learners to grasp the actual meaning of the words before beginning a task.

Though the provision of more time and pre-planning does allow for a better-written piece overall, it does not explicitly imply that time-pressured writing is redundant or useless. Moreover, the context, content, and purpose of the test may be important, for example, time-pressured writing may lack authenticity and would not be suitable for certain purposes such as those requiring detailed and more accurate representation of a participant's abilities.

REFERENCES

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1), 1–17.
- Caudery, T. (1990). The validity of timed essay tests in the assessment of writing skills. *ELT Journal*, 44, 122-131.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267–296.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367-383.
- Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59–84.
- Hale, G. (1992). *Effects of amount of time allowed on the Test of Written English* (Research Rep. No. 92-27). Princeton, NJ: Educational Testing Service.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30, 461-473.
- Hunt, K. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732–739.
- Knoch, U. & Elders, C. (2009). Validity and fairness implications of varying time conditions on a diagnostic test of academic English writing proficiency. *System*, 38, 63-74. doi:10.1016/j.system.2009.12.006
- Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 140-154). Cambridge: Cambridge University Press.
- Livingston, S. A. (1987). *The effects of time limits on the quality of student-written essays*. Paper presented at the Annual Meeting of the American Educational Research Association, NY.
- Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, 18, 103-118. doi:10.1016/j.jslw.2009.02.003
- Norris, J. & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555-578.

- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- Polio, C., Fleck, C., & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7, 43-68
- Powers, D. E., & Fowles, M. E. (1996). Effects of applying different time limits to a proposed GRE writing test. *Journal of Educational Measurement*, 33, 433-452. doi: 10.1111/j.1745-3984.1996.tb00500.x
- Richards, K. (2003). *Qualitative inquiry in TESOL*. New York, NY: Palgrave Macmillan
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-322). Rowley, MA: Newbury House.
- Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17-30). Oxford, UK: Heinemann
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). New York: Cambridge University Press
- VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287-301.
- West, M. (1974). *A general service list of English words with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London: Longman, 1953.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Technical Report #17. Honolulu, HI: National Foreign Language Resource Center.