# THE EFFECTS OF PLANNING WITH WRITING ON THE FLUENCY, COMPLEXITY, AND ACCURACY OF L2 ORAL NARRATIVES

By

Hiep Thien Chau

# A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Second Language Studies - Doctor of Philosophy

## ABSTRACT

# THE EFFECTS OF PLANNING WITH WRITING ON THE FLUENCY, COMPLEXITY, AND ACCURACY OF L2 ORAL NARRATIVES

By

# Hiep Thien Chau

This study is aimed to investigate whether planning with writing enhances the fluency, complexity, and accuracy of L2 oral narratives. Three groups of intermediate EFL learners at a university in Vietnam with 30 participants in each performed a picture-based narrative under one of the conditions: no planning, planning without writing (rehearsal), and planning with writing. Given 10 minutes of planning, the planning-without-writing group were told to rehearse their performance while the planning-with-writing group were told to write out the narrative in full sentences. A post-task interview was also conducted to probe what participants chose to attend to while planning.

All the oral performances were transcribed and analyzed using a comprehensive set of measures for fluency, complexity, and accuracy. One-way ANOVA results showed that both rehearsal and writing before speaking had significant effects on all the three aspects of oral production, but there was no significant difference between planning with and without writing. The rehearsals and written narratives during planning, which were analyzed and triangulated with the interview data, also revealed that both planning groups, in general, had similar patterns of planning with lexical search taking most of their time. The planning-with-writing group tended to focus more on form than the planning-without-writing group.

The evidence from this study furthers our understanding of the effect of planning on oral task-based performance and supports a comprehensive approach to detecting the planning effect

using both general and specific measures. The findings also inform L2 researchers and teachers of the relationship between writing and speaking in L2 development.

To my mother and Chris

## ACKNOWLEDGEMENTS

I owe my sincere gratitude to many people for their assistance of various kinds with this project.

First and foremost, I would like to extend special thanks to my chair, Dr. Charlene Polio, for her inspiring me with the dissertation topic and patiently guiding me through the process of conducting the study. I would like to thank my committee members, Dr. Debra Hardison, Dr. Susan Gass, and Dr. Aline Godfroid, for their thought-provoking comments that enlightened me in many aspects.

I would like to thank Dr. Rebecca Foote, Dr. Shawn Loewen, and Dr. Debra Friedman for their wholehearted instruction during the first hard years of my life as a Ph.D. student. Their lessons about quantitative and qualitative research are of great help not only to my dissertation but also to my research life later.

I am eternally grateful to Dr. Christopher Wheeler, Dr. Karen Klomparens, Dr. Jeffrey Riedinger, and Dr. Janet Swenson, who invested time and effort in bringing me to Michigan State University and helping me with the assistantships as well as fellowships. Without their generous support, I would not have been able to go this far.

No word can express my thanks to Dr. Phu Nguyen for everything he did for me as a wise mentor, a dear uncle, and a close friend.

I will not allow myself to forget my honest and helpful friends at Michigan State University. They are Scott Chiu, Ching-Ni Hsieh, Amy Thompson, Luke Plonsky, Maren Schierloh, Minh Duong, Yen Duong, Ngoc Lan Dang, Thanh Ha Nguyen, Uthane Supatti, and many others. They were always willing to share with me many things from useful readings and research experiences to funny stories and even unbearable frustrations.

v

I am also indebted to all of my colleagues and students at Cantho University for their willingness to assist in recruiting the participants and collecting data for my research. I wish I could hug each of them and tell them how much their help meant to me.

Finally, without the love of my wife Bao Dien and two sons Tuc and Khiet, I could not have had enough stamina to complete this dissertation.

LIST OF TABLES	X
LIST OF FIGURES	xi
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	6
Models of first and second language speech production	6
Measuring language production	8
Planning research and issues	10
Trade-off effects	13
Task complexity	14
Guided planning	15
Planning with writing	17
The present study: Research questions and hypotheses	20
CHAPTER 3: METHODOLOGY	23
Research design	23
Participants	23
Materials	25
Pretests	25
Oral task	27
No planning (NP)	29
Planning without writing (P-W)	29
Planning with writing (P+W)	29
Interview	30
Procedure	30
Measures	33
Fluency measures	33
Speed fluency	33
Breakdown fluency	34
Repair fluency	34
Complexity measures	35
Overall complexity	35
Phrasal complexity	35
Subordination	36
Structural variety	36
Lexical variety	36
Lexical density	37
Accuracy measures	37
Errors per 100 words	37
Correct verb forms	37
Target-like use (TLU) of articles	

Target-like use of the plural -s	38
Lexical errors	38
Analysis	39
Analysis of pretests	39
Transcribing oral performances, rehearsals, and interviews	39
Analysis of oral performances	40
Analysis of planning data	43
Analysis of interview data	44
CHAPTER 4: RESULTS	45
RQ 1: What effects does planning with and without writing have on the fluency of a	n L2
oral narrative?	45
Speed Fluency	45
Breakdown Fluency	46
Repair Fluency	50
RQ 2: What effects does planning with and without writing have on the complexity of	of an
L2 oral narrative?	53
RQ 3: What effects does planning with and without writing have on the accuracy of	an
L2 oral narrative?	57
RQ 4: What do learners attend to during planning?	60
Results from analysis of self-repairs in the planning data	61
Results from qualitative analysis of the interview data	62
Interview question 1: "What were you thinking about while planning"	?" .63
Attention to content	64
Attention to vocabulary	65
Attention to grammar	65
Attention to pronunciation	66
Interview question 2: "What were you most concerned about during	
planning?"	68
Interview question 3: "Do you think you performed the oral task	
successfully?"	69
Interview question 4: "Could you use what you had planned in your	
narrative?"	72
Interview question 5: "What are the benefits of the type of planning y	ou
experienced?"	74
Interview question 6: "If you are allowed some time to plan for anoth	er
similar oral task in the future, how would you choose to plan your	
speaking?"	77
Results from qualitative analysis of the four special cases	78
CHAPTER 5: DISCUSSION AND CONCLUSION	86
RQ 1: What effects does planning with and without writing have on the fluency of an	n L2
oral narrative?	86
RQ 2: What effects does planning with and without writing have on the complexity of	of an
L2 oral narrative?	88

RQ3: What effects does planning with and without writing have on the accuracy	of an L2
oral narrative?	90
RQ4: What do learners attend to during planning?	91
Limitations of the study and suggestions for future research	93
Contributions of the study	
APPENDICES	96
Appendix A: Grammar pretest	97
Appendix B: Pronunciation pretest	
Appendix C: The picture set for the task	105
Appendix D: Task instructions in English and in Vietnamese	
Appendix E: Interview questions	
Appendix F: Coding guidelines for fluency measures	110
Appendix G: Coding guidelines for complexity measures	111
Appendix H: Coding guidelines for accuracy measures	112
Appendix I: Data samples	116
Appendix J: Inter-rater reliability	118
REFERENCES	119

# LIST OF TABLES

Table 1: Distribution of participants by gender across three groups
Table 2: Descriptive statistics for pronunciation and grammar pretest scores by group
Table 3: Descriptive statistics for speed fluency measure
Table 4: One-way ANOVA results for speed fluency measure
Table 5: Descriptive statistics for breakdown fluency measures
Table 6: One-way ANOVA results for breakdown fluency measures
Table 7: Tukey HSD comparisons for breakdown fluency measures  49
Table 8: Descriptive statistics for repair fluency measures  51
Table 9: Mean rankings for repair fluency measures  52
Table 10: Kruskal-Wallis test results for repair fluency measures  52
Table 11: Descriptive statistics for complexity measures
Table 12: One-way ANOVA results for complexity measures  54
Table 13: Tukey HSD comparisons for complexity measures
Table 14: Descriptive statistics for accuracy measures
Table 15: One-way ANOVA results for accuracy measures  59
Table 16: Tukey HSD comparisons for accuracy measures
Table 17: Coding guidelines for fluency measures  110
Table 18: Coding guidelines for complexity measures  111
Table 19: Coding guidelines for accuracy measures  112
Table 20: Inter-rater reliability  118

# LIST OF FIGURES

Figure 1: The picture set for the tasl	
--	--

## **CHAPTER 1: INTRODUCTION**

"Are you going to have an oral dissertation defense?"

"Yes, I am terribly worried."

"Don't panic. Be well-prepared."

"Thank you. I know, but how?"

The conversation above may reflect part of a real-life scenario in which one speaker has an important oral task and needs to prepare for it. Normally, the more important and complex the task is, the more time and effort the speaker has to utilize for preparation. The process of planning for speaking may involve multiple rehearsals, making notes, creating an outline, or even writing out the whole script with the simple aim of improving oral fluency during task performance. These common planning activities can be easily observed not only in real life but also in second language (L2) learning because L2 learners need more time than L1 speakers to retrieve appropriate vocabulary and plan the structures for expressing their message as well as try to articulate their speech. In other words, the process of formulating the language representation and articulating it in L2 speech production is less automatic than in L1 production (Kormos, 2011). As a result, pre-task planning is essential and beneficial for L2 learners' interlanguage development, especially in the context of foreign language teaching and learning where learners have few opportunities to practice using the target language outside the classroom.

As in most of the foreign language environments, adult English as a foreign language (EFL) learners in Vietnam have more exposure to written English than spoken English. Through seven years of lower and upper secondary school, students have no more than two hours and a

half per week for studying English and have to constantly prepare for written English tests that focus on reading, grammar, and sentence transformation. This heavily grammar-based tradition comes from the application of the Grammar-Translation Method because many EFL teachers see it as suitable for dealing with large classes in Vietnam while they have to teach under the curriculum pressure, with limited time, and with communication in English not being the top priority of the formal language instruction.

Another reason for the dominance of the Grammar-Translation Method in the Vietnam context is that teachers tend to teach in the way they have been taught, and many generations of EFL teachers in Vietnam have learned English mainly by reading, doing grammar exercises, and taking numerous structure-based tests. In fact, Vietnamese learners of English learn the foreign language for testing, not for acquisition. Even when more innovative methods and approaches for communication in class are introduced at college level, they have been struggling to find a place in a teaching and learning culture where classrooms are mostly teacher-centered and the noise of students practicing communication in class is sometimes unacceptable to many other teachers teaching nearby. As a result of this teaching practice, it is not surprising that Vietnamese EFL students tend to be better at writing than speaking English and often rely on their explicit knowledge during language production. Experienced in teaching in such a context, I have been wondering if EFL students can carry the vocabulary and structures in their writing over to speaking and if English teachers can help students succeed in doing so by asking them to plan with writing before speaking.

Over the last two decades, the impact of pre-task planning on second language production has attracted extensive research due to its practical benefits in the task-based language teaching program (e.g., Ellis, 1987; Foster & Skehan, 1996; Kawauchi, 2005; Mochizuki & Ortega, 2008;

Yuan & Ellis, 2003). The effects of planning have been examined mainly based on three aspects of language production: fluency, complexity, and accuracy (Skehan, 1998; Skehan & Foster, 1999). There is a consensus that pre-task planning has a beneficial effect on fluency, but there are mixed results for complexity and accuracy (Ellis, 2005, 2009). A question that remains unanswered is why planning seems not to enhance all the three aspects of language production simultaneously.

The problem with mixed results in pre-task planning research has fueled a debate between the advocates of the Limited Capacity Hypothesis (e.g., Skehan, 1998; Bygate, 1999) and the Cognition Hypothesis (Robinson, 2001a, 2005, 2011). The Limited Capacity Hypothesis claims that humans possess limited attentional resources, thus they cannot attend to many aspects of language under time pressure, resulting in a trade-off between them, which means learners may focus on one of the three dimensions—fluency, complexity, and accuracy--to the detriment of the other two. This trade-off effect was assumed to explain why previous studies on pre-task planning failed to prove the parallel gains in the three components of language production.

However, the Cognition Hypothesis claims that learners can simultaneously access multiple and non-competitional attentional pools and that complex tasks should promote more accurate and complex, though less fluent, language than simpler counterpart tasks. As a consequence, if task complexity is manipulated by increasing the cognitive demands of a task, complexity and accuracy can be improved at the same time. If Robinson is right, the problem that needs to be solved is how researchers can design a task to be cognitively complex enough to orient learner attention to the linguistic aspects of the output. In addition, as categorized in the Triadic Componential Framework developed by Robinson (2007), planning time is a task complexity characteristic that can support L2 learners in producing more complex and accurate

language if the task is manipulated as a cognitively complex task. Little research to date has provided convincing evidence for the comprehensive benefits of pre-task planning, thus partly explained why the debate over the role of planning time between the two hypotheses has not been settled yet.

Moreover, it is possible that whether or not a task is cognitively demanding chiefly depends on how researchers or teachers instruct the learners to perform the task. There has been a paucity of studies looking at the role of detailed guidance (Mochizuki & Ortega, 2008; Sagarun, 2005) in planning and performing a task. Thus, there is also a need to reanalyze the guidance for pre-task planning in previous research to see how it influences learner attention during planning and performance. Another important question is whether the mode of planning can enhance learner focus on form. While writing appears to direct learners to focus more on form (Ellis, 1987; Weissberg, 2000; Williams, 2008), few studies have investigated writing as a planning activity prior to oral performance (Kawauchi, 2005).

Motivated by the aforementioned gaps, the present study is intended to investigate whether pre-task planning with writing can improve the fluency, complexity, and accuracy of L2 oral narratives. The results are expected to satisfy the hypotheses that planning with writing will induce greater fluency, complexity, and accuracy in the L2 learners' oral narratives than no planning and that planning with writing will enable L2 learners to attend more to grammatical form than no planning.

The current study contributes to our knowledge by addressing important issues in pretask planning studies. First, the findings of the study provide evidence for the supporting role of planning as a task characteristic that can help develop the fluency, complexity, and accuracy of L2 production as proposed by the Cognition Hypothesis. Second, I argue that the role of task

instructions and detailed guidance for planning are of great significance in deciding the complexity of a task. Third, given the comprehensive measures of fluency, complexity, and accuracy in the study, researchers may have confidence in designing more task-specific or sensitive measures to gauge the effects. Fourth, the effect of planning with writing on oral production gives insights into the writing-speaking connections. Ultimately, the study suggests some pedagogical implications for the task-based teaching programs.

This dissertation is organized in five chapters. Chapter 1 presents an overview of the rationale, purpose, context, and significance of the study. Chapter 2 reviews the relevant literature and previous studies on pre-task planning. Chapter 3 explains the methodology of the study. Chapter 4 reports the results of quantitative and qualitative analyses. Finally, Chapter 5 discusses the findings and their implications, acknowledges the limitations of the study, and evaluates the contributions to the body of pre-task planning research.

### **CHAPTER 2: LITERATURE REVIEW**

In this chapter, I first review the models of first and second language speech production as theoretical background for understanding the processing mechanisms involved in performing oral tasks. This review will be followed by an examination of the constructs commonly used to measure language production, namely *fluency*, *complexity*, and *accuracy*. Next, I present previous planning studies and discuss their confounding results from the perspectives of the Limited Capacity Hypothesis (e.g., Skehan, 1998; Bygate, 1999) and the Cognition Hypothesis (Robinson, 2001a, 2005, 2011). Then, I propose manipulating detailed guidance for planning and writing as a planning activity to promote more focus on form during planning and performance. The chapter ends with the research questions and hypotheses of the study.

## Models of first and second language speech production

The most influential model of first language (L1) speech production in psycholinguistics is probably the one developed by Levelt (1989, 1999). This model consists of three main stages, namely, the *conceptualizer*, the *formulator*, and the *articulator*, processing language in a unidirectional, incremental way. In the stage of conceptualization, language production starts with the processes of conceiving an intention to express, selecting the relevant information from long-term memory or the environment, segmenting and ordering that information for constructing the intended utterances. Levelt also distinguished between *macro-planning* and *micro-planning* conceptualization processes. Macro-planning involves the elaboration of a communicative goal expressed by speech acts (e.g., informing, directing, requesting, and so on) and the retrieval of appropriate information. Micro-planning involves assigning the right propositional shape to these chunks of information and deciding on matters such as what the

topic or focus of the utterance will be. The product of conceptualization is called the preverbal message with all the essential information to transform meaning into language. Following conceptualization is the stage of formulation responsible for the grammatical, lexical, and phonological encoding of the message. It means that at this stage speakers have to translate the conceptual representation into a linguistic form by selecting the individual words that they want to say and put them together to form a sentence. Levelt considered this stage "a largely automatic process" (p.21). Finally, the processes of articulation involve detailed phonetic and articulatory planning to ensure that the sounds must be produced in the correct sequence in overt speech. In addition, according to Levelt, the language produced is assumed to be monitored at all three levels. The process of monitoring includes checking the correctness and appropriateness of the speech production. It can be inferred that if a planning opportunity is given to task performers, they can use it to conduct monitoring at any of the above levels.

Extending Levelt's (1989, 1999) work as well as incorporating previous models of L2 speech production (de Bot, 1992; Poulisse & Bongaerts, 1994; Towell et al., 1996), Kormos (2011) proposed a bilingual speech production. She first noted that there is a consensus among speech production researchers on four important components of language production: conceptualization, formulation, articulation, and self-monitoring, among which conceptualization, formulation, and articulation follow each other in this order. Additionally, in L1 production, planning the message requires attention, whereas formulation and articulation are mainly automatic, and therefore processing mechanisms can work in parallel, making L1 speech generally smooth and fast. Nevertheless, for bilingual speakers, many of the syntactic and phonological rules in L2 are not automatized and are assumed to be stored in a declarative memory of L2 specific knowledge. This is the fourth knowledge store that Kormos (2011) added

to the three other knowledge stores proposed in Levelt's (1999) model which include the store for the knowledge of the external and internal world, the mental lexicon, and the syllabary. She finally acknowledged that despite a number of differences existing between first and second language speech production, the basic psycholinguistic mechanisms in producing speech seem to be very similar (Kormos, 2006b, 2011).

Segalowitz (2010) also proposed a model of the L2 speaker, adapted from Levelt's (1999) "blueprint" of the monolingual speaker, to which he added fluency vulnerability points where L2 speakers' disfluencies could result from their difficulties in speech processing. These critical points include microplanning, grammatical encoding, lexical retrieval, morphophonological encoding, phonetic encoding, articulation, and self-perception. Perhaps, planning time could be conceptualized as a way for L2 speakers to circumvent these vulnerability points.

## Measuring language production

Many researchers believe that the constructs of L2 performance are multi-dimensional in nature and that these dimensions can be comprehensively captured by the notions of *fluency*, *complexity* and *accuracy* (Ellis, 2003; Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Norris & Ortega, 2009; Skehan, 1998). Fluency "concerns the learner's capacity to produce language in real time without undue pausing or hesitation" (Skehan, 1996, p. 22). Various ways of measuring this have been devised--speech rate, pause length, silence, false starts, repetitions, and reformulations. This interpretation of fluency is distinct from the concept of fluency as overall proficiency (Derwing et al., 2009; Lennon, 1990; Skehan & Foster, 1999; Trofimovich & Baker, 2006). Furthermore, it is important to note that though fluency denotes an automatic procedural skill on the part of the speaker, it can also be considered from the standpoint of the listener

(Derwing et al., 2009). Previous studies of oral fluency found that fillers, excessive and inappropriate pausing, false starts, self-corrections, and a slow speech rate can all negatively affect native listeners' judgments on L2 comprehensibility (Derwing & Munro, 2001; Derwing et al., 2007; Munro & Derwing, 2001). Therefore, given the unavailability of listener judgments, researchers can treat fluency as a multi-dimensional construct consisting of three components: speed fluency (e.g., number of syllables per minute), breakdown fluency (e.g., filled and unfilled pauses per minute), and repair fluency (e.g., self-corrections, repetitions, replacements, and false starts per minute) (Housen & Kuiken, 2009; Skehan, 2009; Tavakoli & Skehan, 2005).

As for complexity, this term was first introduced in an L2 model by Skehan (1989) which included fluency, complexity, and accuracy as the three principal proficiency dimensions. Complexity concerns "the extent to which the language produced in performing a task is elaborate and varied" (Ellis, 2003, p. 340). This definition of complexity refers to the linguistic complexity of L2 performance and proficiency. According to Housen and Kuiken (2009), linguistic complexity has been commonly understood as "the size, elaborateness, richness, and *diversity* of the learner's linguistic L2 system" (p.464). Measures of complexity are generally based on the extent to which subordination is evident (e.g., number of clauses per T-unit or Cunit) (Elder & Iwashita, 2005; Ellis & Yuan, 2005; Kawauchi, 2005). In some studies, lexical complexity has been assessed (e.g., by means of type-token ratio) (Laufer, 1991; Ortega, 1999; Robinson, 2001b). However, as Norris and Ortega (2009) pointed out, the way researchers operationalize the construct of complexity has been impoverished. They suggest using more specific, dynamic, and sensitive measures in addition to general measures in order to tap complexity multidimensionally such as structural variety and phrasal elaboration (e.g., mean length of clause).

Regarding accuracy, Housen and Kuiken (2009) claim that it is "probably the oldest, most transparent and most consistent construct" (p. 463). Accuracy refers to the extent to which the language produced conforms to target language norms (Wolfe-Quintero et al., 1998), thus deviations from the norm are usually characterized as errors. Researchers have varied in how they measured accuracy. Some (e.g., Crookes, 1989; Wigglesworth, 1997) have preferred to examine how accurately specific grammatical features (e.g., articles) are used while others have elected to use more generalized measures such as percentage of error-free clauses (Skehan & Foster, 2005; Yuan & Ellis, 2003). To better measure the accuracy of L2 performance as a multicomponential construct like fluency and complexity, both general and task-specific measures should be employed as proposed by Housen and Kuiken (2009) as well as Norris and Ortega (2009).

According to Skehan (1998), these three aspects of performance need to be distinguished because they are differentially affected by the particular type of processing a learner adopts. He suggests that under certain conditions learners will choose to draw on their lexicalized knowledge of language, resulting in enhanced fluency, while under others they will be able to refer to their rule-based system, leading to greater complexity and/or accuracy. Among the conditions that Skehan identifies as influential in this respect is the opportunity for planning.

#### Planning research and issues

Over the last 20 years, studies have investigated the impact of planning on language production (e.g., Crookes, 1989; Ellis, 1987; Foster & Skehan, 1996; Kawauchi, 2005; Mehnert, 1998; Mochizuki & Ortega, 2008; Ortega, 1999; Sangarun, 2005; Yuan & Ellis, 2003). Most studies are based on information processing theory, which claims that humans possess a limited

processing capacity, and thus are not able to fully attend to all aspects of a task simultaneously (Anderson, 1995; Baddeley & Logie, 1999). Second language (L2) learners, especially those with limited proficiency, seem to find it difficult to attend to meaning and form at the same time, so they have to make decisions about how to allocate their attentional resources by prioritizing one aspect of language over others (Anderson, 1995; Skehan, 1996, 1998; Skehan & Foster, 1999, 2001; VanPatten, 1990). Thus, it is believed that L2 learners can compensate for these processing limitations if they have an opportunity to plan the linguistic and propositional content of an upcoming task; as a result, the quality of their linguistic output could improve (Skehan, 1996).

Research has so far supported the claim that planning in advance has a beneficial effect on fluency, but the results for complexity and accuracy are more mixed (Ellis, 2009; Ortega, 1999). Pre-task planning has shown to result in increased fluency (e.g., Crookes, 1989; Foster & Skehan, 1996; Gilabert, 2007; Skehan & Foster, 2005; Tavokoli & Skehan, 2005; Wigglesworth, 1997; Yuan & Ellis, 2003). Mehnert (1998) investigated different lengths of planning time (none, 1 minute, 5 minutes, and 10 minutes) and found that oral fluency improved with each increase in planning time. Interestingly, planning appears to have a greater effect on fluency in the case of less complex tasks; for instance, Skehan and Foster (1997) reported that planners paused less than non-planners in the personal information and narrative tasks but not in the decision-making task.

Many studies also report a positive effect of planning time on the complexity of learners' oral productions (e.g., Ortega, 1999; Sangarun, 2005; Tavokoli & Skehan, 2005; Wigglesworth, 1997; Yuan & Ellis, 2003); however, the results are not consistent. Those studies that involved a monologic performance (Kawauchi, 2005; Yuan & Ellis, 2003) reported that planning assisted

complexity. In relation to task complexity, Foster and Skehan (1996) found that pre-task planning led to greater grammatical richness in their personal information and narrative tasks but not in the decision-making task; in contrast, Skehan and Foster (1997) failed to find any effect of planning on complexity in the narrative task. Regarding planning variables, Mehnert's (1998) study found no effect for planning on complexity in any of her time conditions. Neither Kawauchi (2005) nor Mochizuki and Ortega (2008) found any significant difference in the effect of the different types of planning they investigated. However, there is some evidence that the degree of guidance provided influences outcomes. Foster and Skehan (1996) found that detailed planning resulted in greater grammatical complexity for the first 5 minutes of task performance than undetailed planning, a finding that they replicated in Skehan and Foster (2005).

More mixed results have been reported for accuracy. Previous studies suggest that a number of factors influence whether pre-task planning leads to more accuracy. One factor that appears to influence the results for accuracy is the learner's proficiency. Kawauchi (2005) found that planning had much more effect with learners of low proficiency than with advanced level learners. Specifically, she reported that only the low-proficiency learners improved on regular past tense as a result of the planning opportunity. This study suggests the importance of controlling for learner proficiency when investigating the effects of planning on accuracy. Regarding planning time, no significant difference was found in linguistic accuracy between groups that were given planning time and those that were not (Wigglesworth, 1997; Yuan and Ellis, 2003). However, Mehnert (1998) found that increased accuracy was evident with learners who were given just 1 minute to plan but that allowing more time (5 minutes or 10 minutes) did not result in any additional gains in accuracy. Research findings for accuracy also vary from one linguistic feature to another. Ellis (1987) found that planning had a positive effect on the

accuracy of regular, "rule-governed" past tense forms (e.g., *worked*), but not on the accuracy of irregular past tense forms (e.g., *went*). However, Crookes (1989) found no effect on the use of articles. In her study of Spanish learners, Ortega (1999) found positive accuracy effects for planning on noun-modifiers but not on articles. Additionally, Skehan and Foster (1997) found that the type of task influenced whether planning had an effect on accuracy; planning led to increased accuracy in the case of a personal and a narrative task, but not in a decision-making task. In another joint study, Foster and Skehan (1996) showed that pre-task planning had an effect on general linguistic accuracy when planning was unguided, but not when it was guided. Nevertheless, the evidence in support of some of these generalizations remains meager.

# Trade-off effects

To explain why the effects of planning on complexity and accuracy are more variable, researchers who hold the view that humans have limited attention mechanism and processing capacity (e.g., Skehan, 1998; Bygate, 1999; Skehan & Foster, 1997, 1999) proposed a trade-off effect. In their view, fluency is seen as an aspect of L2 production competing for attentional resources with accuracy, while accuracy in turn competes with complexity. Learners may focus on one of the three dimensions to the detriment of the other two. For instance, Skehan and Foster (1997) argued that the planners in their study were able to use the planning time to attend to accuracy because they did not need to devote much attention to encoding the content in the picture-based narrative task. On the other hand, in the decision-making task, which was inherently unstructured, they used the planning time to sort out how to express complex ideas; as a result, little capacity was left to attend to accuracy.

## Task complexity

A different view known as the Cognition Hypothesis proposed by Robinson (2001a, 2001b, 2003, 2005, 2011) claims that learners can simultaneously access multiple and noncompetitional attentional pools and that complex tasks should promote more accurate and complex, though less fluent, language than simpler counterpart tasks. More specifically, Robinson distinguishes between two subgroups of task variables contributing to task complexity -resource-directing and resource-dispersing dimensions-to explain task effects. Resourcedirecting dimensions (e.g., requiring temporal reference, and/or spatial reasoning) make conceptual/communicative demands; thus, increasing task complexity along these dimensions has the potential to direct learners' attentional and memory resources to lexical, morphological, and syntactic aspects of the L2 system required to accurately understand and convey concepts, such as space and time as well as motion. For instance, tasks requiring reference to *time* should promote the use of more developmentally advanced L2 tense and aspectual encoding (Shirai, 2002 as cited in Robinson, Cadierno, & Shirai, 2009). In contrast, resource-dispersing dimensions (e.g., including planning time, task structure, and prior knowledge) place performative/procedural demands on cognition, thus facilitating automatic access to and control of existing interlanguage resources. However, increasing complexity along these dimensions (e.g., by removing planning time) does not direct learner attention to features of linguistic code but simply disperses it over many linguistic and non-linguistic aspects of the task. On this resource-dispersing dimension of task complexity, the cognition hypothesis predicts lower accuracy, complexity, and fluency of performance for tasks performed with less support available, compared to those where support is provided. In such cases, the effects of increasing the complexity of a resource-directing characteristic (e.g., by requiring temporal reference) may

be weakened or negated by increasing the complexity of a resource-dispersing characteristic (e.g., by taking away planning time). Thus, according to the Cognition Hypothesis, there are likely to be "synergic effects" on speech production when tasks are made complex along both resource-directing and resource-dispersing dimensions simultaneously (Robinson, 2011, p. 21).

There have been several studies presenting data supportive of the Cognition Hypothesis (Gilabert, 2007; Ishikawa, 2007; Iwashita, McNamara, & Elder, 2001; Kuiken & Vedder, 2008; Robinson, 1995). However, as Skehan (2009) points out, the evidence is still limited because some studies report raised accuracy, but no increase in complexity (Gilabert, 2007; Kuiken & Veddar, 2008), and some others involve written, not oral performance (Ishikawa, 2007; Kuiken & Vedder, 2008). A recent meta-analysis of nine comparable studies manipulating task complexity along the resource-directing dimensions revealed small positive effects for accuracy during monologic task performances and small negative effects for fluency, which is concluded to be consistent with predictions of the Cognition Hypothesis (Jackson & Suethanapornkul, 2013). Nevertheless, predictions about syntactic complexity are not supported by the findings of the study.

Therefore, the present study is another attempt in this strand of research in the hope of achieving more balanced gains in all the three aspects of L2 oral performance—complexity, accuracy, and fluency--by manipulating the mediating role of some combined factors such as planning time, detailed guidance, and writing as a planning activity.

## Guided planning

An interesting strategy for investigating how to promote the quality of language production is "to manipulate the very nature of planning" by modifying instructions for planning

(Mochizuki & Ortega, 2008, p. 15). Indeed, instructions before planning play a very critical role in planning activities since they may focus the learners' attention on a certain aspect of language performance rather than another. Consequently, a particular concern has been with whether this strategy can foster linguistic accuracy, an area in which planning effects have been inconclusive. The reason for not being able to see the effect of planning on accuracy in L2 production may have been that the instructions provided in previous studies were too broad (e.g., asking the learners to consider "the syntax, lexis, content, and organization of what they would say") (Foster & Skehan, 1996) or rather vague (e.g., "try to think of the vocabulary and grammar you may use in the story") (Yuan & Ellis, 2003). Although Foster and Skehan (1996) hypothesized that detailed/guided planning would help learners achieve greater accuracy during oral production than undetailed/unguided planning and no planning, they found no difference possibly because what they called detailed/guided planning may not have been focused enough to direct learners' attention to features of linguistic code.

In the strand of guided planning research, Sangarun (2005) asked 40 Thai EFL students to perform oral tasks under one of four conditions: minimal, meaning-focused, form-focused, and meaning-plus-form-focused planning. In the form-focused condition, students were asked to pay attention to up to four specific structures in each task. Unlike Foster and Skehan (1996), Sangarun found positive effects for the three planning conditions on the speech quality, including accuracy. She argued that the types of planning in previous studies were not operationalized in a way that could encourage attention to form. Most recently, Mochizuki and Ortega (2008) succeeded in getting improved accuracy by providing specific instructions on language to be used with a brief handout about how to use relative clauses, resulting in more accurate relative clauses in the narratives of the guided planning group while showing the same global levels of

complexity and fluency in comparison with the no planning and unguided planning groups. Their results indicate that greater accuracy in language performance may be achieved if guided planning is defined as a condition in which planners are given specific grammatical guidance to focus their attention on certain linguistic forms. Nevertheless, the definition of guided planning has been ambiguous since it is unknown whether it should guide learners what to focus on during planning or how to plan their performance.

# Planning with writing

There have been legitimate arguments for linking speaking and writing in L2 classrooms (e.g., Weissberg, 2006). However, while the role of speaking in scaffolding writing is increasingly well-supported empirically through research on peer review, student-teacher conferencing, collaborative talk, or tutoring (e.g., Liu & Hansen, 2002; Pathey-Chavez & Ferris, 1997; Storch, 1999; Williams, 2004), research on how writing can support speaking has received only modest attention. It is argued that literacy plays a central role in second language acquisition in classroom settings (Harklau, 2002) and particularly has important effects on human oral language processing (Tarone & Bigelow, 2005). Writing obviously provides a means of practicing the language one is learning to speak, and the act of reflection during writing is theoretically claimed to slow down cognitive processing of language, offering great opportunities for planning and allowing for explicit attention, or noticing (Williams, 2008).

As Williams (2008) notes, research has indicated not only that learners introduce forms in their writing that they have not yet used in speaking, but also that using a form in the planned context of writing may lead to use in unplanned speaking contexts. For example, Weissberg's (2000) study, comparing the acquisition of syntactic features in the speech and writing of five L2

learners of English, found that writing (in paired written tasks) appeared to be "the preferred medium for the emergence of new morpho-syntactic forms" (p. 37). In addition, he found that learners' use of these forms was more accurate in writing than in speech, a finding in line with previous research on planning time in oral production (Ellis, 1987). A longitudinal study (Kim, 2008) of two beginning ESL students enrolled in kindergarten also showed that integrated instruction (i.e., oral plus written) led to greater gains in the students' oral language development than did an exclusively oral-only intervention. Similarly, in another study of 44 ESL learners, Adams and Ross-Feldman (2008) have observed that collaborative writing and speaking does appear to encourage more learner attention to form measured through the use of language-related episodes (LREs) than collaborative speaking only. However, the measures of this study did not reach statistical significance possibly due to small sample size; in addition, the task design did not require writing before engagement in collaborative speaking, so the influence of writing on speaking cannot be determined yet and thus needs further research.

Recently, based on his investigation of 34 ESL learners, Blake (2009) argued for textbased Internet chatting as a helpful way of improving oral fluency because the automatization of lexical and grammatical knowledge can be facilitated as the learners formulate their message. Hardison's (2011) study has also found that writing in the form of various types of electronic communication (versus writing course papers and other forms) is a significant predictor of oral fluency for advanced ESL learners in L2 interactions. Fluency in this case was judged by native speaking English raters who listened to the interactions of 124 Korean ESL graduate students. This result is consistent with Payne's (2002) finding that those engaging in online chatting showed greater gains on an oral proficiency test than those who did not, leading him to conclude that there is a transfer of skills from writing to speaking.

Up to now, there has been no study investigating pre-task planning with writing except Ellis' (1987) and Kawauchi's (2005) study. Though Ellis' study was not originally designed for this reason, the experimental conditions in his study were similar to those which involve writing as a planning activity before performing an oral task. Ellis asked all 17 ESL learners to perform a narrative task under three conditions. In the first condition, they were asked to perform a written narrative based on a set of pictures. In the second condition, they orally retold the same story as in condition 1 but without access to their writing already done. The learners were allowed to record the story twice, but only the second retelling was transcribed and analyzed, which means the participants had the opportunity to rehearse their performance. In the third condition, the learners performed another oral narrative based on a set of pictures different from that in condition one. The results of the study showed that the most accurate use of regular past tense was evidenced in condition 1, followed by condition 2, and then by condition 3 despite no statistically significant difference between the conditions 2 and 3. Though Ellis' study confounds planning and modality conditions (i.e. condition 1 involved a written task while conditions 2 and 3 involved oral tasks), writing in condition 1 appeared to afford more opportunity for focus on form.

Unlike Ellis' study, Kawauchi's (2005) study involved writing as a planning activity for several reasons. First, writing may help the researcher to specify the content of planning. Most of the previous studies have attempted to examine planning effects by requiring note-making, a form of writing, to ensure that they were engaged in planning. Usually, what learners did during the planning time was often left to themselves, so it would not be clear what they were actually doing while planning. If learners had been asked to write, we would have had a better picture of their planning content. It is also observable that when people are given time to plan their speech

and are left to their own devices, many of them frequently write it out. More importantly, writing during planning is a kind of pushed output, as in Swain's output hypothesis which claims that learners need to be pushed to make use of their resources, and then upon reflecting on their output they need to consider ways of modifying it to enhance comprehensibility, appropriateness and accuracy (Swain, 2005). However, Kawauchi's (2005) study found no significant difference between the effects of planning with writing and those of the other planning conditions like rehearsal and reading on the fluency, complexity, and accuracy of oral narratives.

# The present study: Research questions and hypotheses

Given a strong need for more evidence to confirm the effects of pre-task planning on language accuracy and complexity in addition to fluency, further research has to be conducted to find a favorable planning condition in which all the three aspects of linguistic performance can be enhanced. I am particularly interested in two types of planning: planning with writing and planning without writing (rehearsal), taking into account such factors as task complexity, detailed guidance of how to plan, and writing as a planning activity. Therefore, the current study is designed to investigate the effect of planning opportunity on the fluency, complexity, and accuracy of oral performance under three conditions: planning with writing, planning without writing (rehearsal), and no planning. Specifically, it addresses the following questions and hypotheses:

*Research Question 1* (RQ1): What effects does planning with and without writing have on the *fluency* of an L2 oral narrative?

*Hypothesis 1*: Planning with and without writing will induce greater fluency in the L2 learners' oral narratives than no planning, as predicted by the Cognition Hypothesis (Robinson, 2001, 2005, 2007, 2011) due to the synergic effects and by pre-task planning studies to date (Ellis, 2009; Ortega, 1999).

*Research Question 2* (RQ2): What effects does planning with and without writing have on the *complexity* of an L2 oral narrative?

*Hypothesis 2*: Planning with and without writing may induce greater complexity in the L2 learners' oral narratives than no planning due to the synergic effects as predicted by the Cognition Hypothesis (Robinson, 2001, 2005, 2007, 2011).

*Hypothesis 3*: Planning with writing may induce greater complexity in the L2 learners' oral narratives than planning without writing and no planning because writing encourages attention to linguistic form and there may be a transfer of skills from writing to speaking, as predicted by Blake (2009), Payne (2002), Weissberg (2000), and Williams (2008).

*Research Question 3* (RQ3): What effects does planning with and without writing have on the *accuracy* of an L2 oral narrative?

*Hypothesis 4*: Planning with and without writing may induce greater accuracy in the L2 learners' oral narratives than no planning due to the synergic effects as predicted by the Cognition Hypothesis (Robinson, 2001, 2005, 2007, 2011).

*Hypothesis 5*: Planning with writing may induce greater accuracy in the L2 learners' oral narratives than planning without writing and no planning because writing encourages attention to

linguistic form and there may be a transfer of skills from writing to speaking, as predicted by Blake (2009), Payne (2002), Weissberg (2000), and Williams (2008).

Research Question 4 (RQ4): What do learners attend to during planning?

*Hypothesis 6*: All the planners will attend to what they view as important (e.g., communication or accuracy) for task performance according to their individual preferences, language expertise, and interpretations of the task demands, as predicted by Ortega (2005).

*Hypothesis* 7: All the planners will attend to more lexical forms than grammatical forms, as predicted by Poole (2005) and Williams (1999).

*Hypothesis* 8: The planning-with-writing group will attend to grammatical forms more than the planning-without-writing and no-planning groups, as predicted by previous research (Adams & Ross-Feldman, 2008; Weissberg, 2000; Williams, 2008).

## **CHAPTER 3: METHODOLOGY**

#### Research design

This study is a single-factor between-subject design with three conditions: *no planning* (NP), *planning without writing* (P-W), and *planning with writing* (P+W). Three groups of participants were asked to orally narrate a story based on a set of six pictures. Prior to performance, two planning groups with and without writing were asked to plan what they were going to say, and some of them were selected for post-task interviews about their planning foci. While the independent variable is the type of planning, the dependent variables encompass both general and specific measures of fluency, complexity, and accuracy for monologic oral production. One-way ANOVAs were conducted to compare the effects of the three conditions.

## Participants

The participants in the study were 90 full-time undergraduate students between the ages of 18 and 21, including 67 females (about 74%) and 23 males (about 26%). They were randomly assigned to one of three experimental groups with 30 participants per group (see Table 1). At the time of data collection, they were mainly second-year and third-year students majoring in English Education, English Studies, and English Translation and Interpreting at a large public university in the Mekong Delta, Vietnam, where the researcher had worked as an instructor. Starting to learn English as a foreign language (EFL) in grade six, they had been exposed to English in formal classroom settings for at least 8-9 years, approximately 2 hours and a half a week in middle and high school (for 7 years) and at least 12 hours a week in their undergraduate programs (for 1-2 years). As English learners in a foreign language environment, these

participants had been exposed to more written than spoken forms of the language and had few opportunities to use the target language for communicative purposes outside the classroom. As they had experienced basic language skills courses for 2 to 4 semesters at the university, their knowledge of written English was expected to be equivalent to the intermediate to high intermediate level. Thus, the participants were assumed to constitute a fairly homogeneous group in terms of their age, educational background, first language, English learning experience, and English proficiency.

Table 1: Distribution of participants by gender across three groups

Experimental Condition						
		NP	P-W	P+W	Total	Percent
Gender	Female	20	25	22	67	74.4
	Male	10	5	8	23	25.6
Total		30	30	30	90	100.0

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planningwith-writing group.

In addition, Skehan (2009), in his re-examination of complexity, accuracy, and fluency (CAF) measures, points to the lack of native speaker data in CAF research. He claims that such data are of crucial importance, as they constitute a baseline along which L2 learners can be compared. For this reason, I recruited 10 native speakers of English (5 males and 5 females) who were undergraduate students at a midwestern U.S. university to serve as a baseline group performing the task with no planning. This group' oral narratives were of great help because they provided a source of data for me to make many error coding decisions, especially decisions on errors in verb tenses, lexical choice, articles, and awkward expressions. However, due to small sample size, the native speakers' data just served as reference data, but were not compared with

the three groups of EFL learners in the present study in terms of fluency, complexity, and accuracy.

### Materials

#### Pretests

To better control for the participants' proficiency level in both written and spoken English, two pretests, one on grammatical structures and the other on pronunciation, were employed. The grammar pretest was a paper-and-pencil test of 80 multiple-choice items constructed to assess the participants' knowledge of important structural and grammatical elements of standard written English at the intermediate level (see Appendix A). This pretest covered a range of grammatical features such as tense and aspect, voice, non-finite verb forms, articles, word order, and subordinate clauses that could be useful in the pre-task planning and performance, especially when writing was involved as a planning activity. Thus, apart from the screening purpose, the grammar test scores were assumed to partly inform the researcher of the existing structures in the participants' interlanguage, and as hypothesized by Ellis (2009), an opportunity to plan is likely to assist the learners in restructuring and developing a better control of their existing L2 knowledge. Therefore, the structures used during planning were expected to carry over to the final narratives with greater accuracy and complexity.

In addition, it is an obvious observation that Vietnamese EFL learners tend to drop the final consonants in speaking English while the participants' ability to produce these sounds would affect the accuracy of the participants' oral performance in the current study. As a result, a short paragraph was prepared for reading aloud to check the participants' pronunciation of the
past tense verbs and adjectives with -ed endings and plural nouns with -s or -es endings (see Appendix B).

The grammar pretest was administered in the classrooms, and the test scores were used to decide which participants would join the second phase of the study in the lab session the following week. The pronunciation test was given right before the experiment in the same lab session, and its results were taken into account in determining whose task performances would be used as part of the analysis. The ninety participants in the present study were those who correctly scored over 40 out of 80 items on the grammar pretest and accurately pronounced over 10 out of 20 targeted words in the paragraph for reading aloud. These average benchmarks were established based upon what was specifically needed for the study to ensure the appropriate and equivalent level of English knowledge and the sample size.

		-		-	95% Confide	ence Interval
Pretest Scores		Ν	М	SD	Lower Bound	Upper Bound
Pronunciation	NP	30	15.03	2.16	14.23	15.84
	P-W	30	15.30	2.51	14.36	16.24
	P+W	30	14.27	2.70	13.26	15.28
	Total	90	14.87	2.48	14.35	15.39
Grammar	NP	30	58.30	7.80	55.39	61.21
	P-W	30	57.10	8.86	53.79	60.41
	P+W	30	58.47	9.81	54.80	62.13
	Total	90	57.96	8.79	56.12	59.80

Table 2: Descriptive statistics for pronunciation and grammar pretest scores by group

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planningwith-writing group. As summarized in Table 2, the descriptive statistics of the participants' pronunciation and grammatical backgrounds did not show much difference between groups. The P+W group had slightly lower pronunciation scores (M = 14.27, SD = 2.70) than the NP (M = 15.03, SD = 2.16) and the P+W (M = 15.30, SD = 2.51), whereas the P-W group had slightly lower grammar scores (M = 57.10, SD = 8.86) than the NP (M = 58.30, SD = 7.80) and the P+W (M = 58.47, SD = 9.81). There was also more variation in pronunciation and grammar scores within the P-W and P+W groups than within the NP.

# Oral task

The participants were required to orally narrate a story based on a series of six pictures taken from Heaton (1975) (see Appendix C), which was the same picture set that Park (2006) and Yuan and Ellis (2003) employed in their pre-task planning studies. Such a common story-telling task in this area of research was thought to facilitate comparison with previous findings. There were no authentic listeners participating in the oral narratives since their presence might encourage many learners "to orient to the listener's needs and to prioritize getting the message across to the listener over being accurate, fluent, or complex", whereas it might make some learners avoid self-corrections and pressure them to prioritize fluency over accuracy (Ortega, 2005, p. 105). Without the presence of listeners, the influence of other interactional variables could also be controlled. In addition, the choice of monologic narratives was justifiable because one of the major concerns of this study was the relationship between writing and speaking. The fact that the oral narratives were monologic and similar to written narratives was expected to advantage the transfer of skills from writing to speaking in the planning-with-writing group.

Regarding task complexity, the oral task in this study was assumed to be conceptually demanding because it required temporal reference and spatial reasoning classified as task

characteristics along resource-directing dimensions by the Cognition Hypothesis (Robinson, 2005, 2011). Also, the participants had to distinguish two groups of boys, two buses, and the time difference on the two clocks in the pictures, whereas they had no prior knowledge of the task type. The piloting of the study revealed that all the participants had little experience in retelling a story based on a set of pictures, and they had great difficulty recognizing the differences between two groups of boys and two buses; as a result, I colored the smaller boys red and the bigger boys blue and highlighted the number of two buses to ensure that the task was reasonably demanding. Additionally, in all the conditions of the study, the participants were allowed to look at the pictures during their oral performance since a recent empirical study (Révész, 2009) on the relationship among tasks, focus-on-form techniques, and L2 development has showed that the group that performed an oral description task with photo support achieved greater gains in using the past progressive form than the group that did not view the photos. She argued that the here-and-now dimension (i.e., referring to here and now or there and then) is resource-directing and the contextual support (e.g., with or without a photo) is resourcedispersing, whereas research to date has operationalized the two dimensions as one variable along resource-directing dimensions. Again, as hypothesized by Robinson (2005, 2011), the impact of task manipulations along resource-directing dimensions can be stronger when the task is simultaneously simpler along one or more resource-dispersing dimensions, so the oral task in this study was designed to benefit from both types of dimensions (i.e., with temporal reference and spatial reasoning in one type and with photo support and/or planning time in the other).

Ninety participants carried out the task under one of the three following conditions (see Appendix D for instructions in English and Vietnamese).

*No planning (NP).* In this condition, the participants were required to perform the task immediately after studying the pictures for a very short time (0.5 minute) so that they would have almost no time for planning in advance.

*Planning without writing (P-W).* For this planning condition, the participants were given 10 minutes to plan prior to task performance as suggested by most previous research (e.g., Crookes, 1989; Kawauchi, 2005; Mehnert, 1998; Yuan & Ellis, 2003) because it would be easier to compare the effects across studies and because when more time (15 minutes) was made available during the pilot study, the participants tended to remember or memorize what they had planned for later recall. The planners in this group were also asked to say out loud what they were preparing in full sentences during planning. The purpose of this practice was to help the researcher identify what the planners noticed rather than how they planned their performance, so I did not employ think-aloud protocols. Encouraging the planners to say everything that came into their minds, which was more like a rehearsal, was expected to mitigate against the tendency to "edit" what is said as can be seen in using think-aloud protocols (Jourdenais, 2001). Moreover, think-aloud techniques may be affected by the different cultural and social perspectives of the participants (Sasaki, 2004) and may change the nature of the activity itself (Jourdenais, 2001); therefore, they were inappropriate for use with Vietnamese students, who tend not to articulate their thoughts, and inapplicable to speaking data that embed a communicative value. No guidance of any other kind in terms of content, organization, or language was provided. The participants were not allowed to write anything, even a few notes.

*Planning with writing* (P+W). This condition also allowed the participants 10 minutes to plan as in the P-W condition. However, the major difference between these two conditions was that under the P+W condition writing out the whole story on a given white sheet was required

during planning, while under the P-W condition the planners rehearsed their performance instead of writing it out. The participants were told that their writing would be taken away after 10 minutes of planning to make sure that the language elicited by the task was oral, thus avoiding the problem in Ellis's (1987) study, where planning and modality were confounded. There might be a time-on-task problem with the P+W group as writing takes longer time. However, if this group had been given more time, they would be spending more time on planning, and it would make the comparison with the P-W group unbalanced.

# Interview

A retrospective semi-structured interview (see Appendix E for the English version) was conducted in Vietnamese right after the experiment to further understand what the participants had chosen to attend to during planning and what benefits as well as difficulties of the planning condition they had experienced. At the end of each experiment session, only one or two participants were interviewed for about ten minutes on a one-on-one basis. The interviewees were the first volunteers that completed the experiment and were willing to share their experience with the researcher. A total of 30 participants with 10 from each group were interviewed in this study. The interview data were used as a source of supplementary information for interpreting the findings of the statistical analyses.

### Procedure

I pilot tested the materials in late spring 2011 with 25 EFL learners and 8 native speakers of English at the Center for Foreign Languages at the intended university in Vietnam. The piloting resulted in modifications of wording and timing in pretests, task instructions, and interview questions. The actual data collection was conducted in fall 2011 in language skills classrooms and a 34-seat language laboratory of the university's English Department, which was considered a safe and familiar environment. All the instructions during the process of data collection were given in Vietnamese.

With the enthusiastic assistance of my colleagues at the institution, I paid the first 45minute visits to 10 intact classes in their usual class sessions to introduce the study, explain the consent form, ask for the class email lists, and administer the grammar pretest. The potential participants were told that all information collected would be kept anonymous and would only be used for research purposes. They were also informed of the bonus credit their instructors would give them for their participation. For instance, the amount of time for the pretests and experiment was equivalent to an hour and a half in the total 30 hours for language skill practice in the classroom in a semester; therefore, a bonus credit equal to 3% of the total grade was given by one of the instructors for completion of the experiment process. For those who did not want to participate in, withdrew from, or partially completed the research for any reasons, the instructors assigned them a written narrative activity that took a similar amount of time so that they could also earn extra credit. The performance of the research began only when the students voluntarily agreed to take the grammar pretest. A total of 178 students completed this test in the classrooms, but only 126 of them (71%), who scored more than 40 correct answers out of 80 items on the test, were chosen for the lab session.

In the following week, I arranged a language lab schedule for the second meeting to record the students' pronunciation, rehearsal, oral task performance, and interviews. I emailed to the 126 students screened by the grammar pretest and had them register for a one-hour session in a three-week schedule. To ensure proper and timely supervision of students' performance, I

scheduled no more than 5 students for each lab session. Those who registered for the same time slots were randomly assigned to the same experimental group.

At the beginning of each lab session, students were seated far enough from the others in order for one's recording not to affect another's. Then, they were asked to fill out a brief information form indicating their identity number assigned in the first meeting by the researcher, ages, semester of study in the program, experimental group number, and number of the assigned computer. Then, they were given a 15-minute tutorial to learn or review how to record their speech and save their audio files on the computer. Familiarizing the students with such voice recording tools as microphones and audio recording software was necessary to minimize their possible anxiety in the real performance. Following the tutorial was pronunciation recording for 5 minutes and the main oral task for 10-20 minutes. When it was time to begin the experiment, I read aloud the instructions in Vietnamese (see Appendix D) and then allowed the students one minute to reread the written instructions by themselves and ask any clarification questions for complete understanding of what they were going to do. According to the experimental conditions, they were required to plan for 10 minutes or not and then perform the task within 5 minutes in as much detail as possible. Since several participants at a time were randomly assigned to the same condition in each session, they started and ended the experiment at the same time. I always reminded the participants to turn on the recording software when they started to tell the story or rehearse their performance.

Finally, right after the oral task performance in each session, I asked the participants if there was anybody interested in a ten-minute interview and willing to share their experience in planning the task, and then chose to interview only one or two students who volunteered first

because of their interest and willingness. I conducted the interviews in Vietnamese and recorded all the responses.

All the sound files (e.g., reading aloud, rehearsal, and oral performance) were taken off the computers in the lab and transferred to the researcher's computer. The writing samples were also collected and put in a folder for later analyses.

For the baseline group of native speakers, I approached 10 American undergraduates individually on the campus of a midwestern university and asked if they were willing to record their task performance without planning on a digital audio recorder. I also had some quick questions about their majors and their comments on the task.

### Measures

Fluency, complicity, and accuracy in this study were addressed as multi-componential constructs, thus both general and specific measures of each dimension were developed to evaluate the quality of the participants' oral production. These measures were dependent variables while the independent variable consisted of the planning type only.

### Fluency measures

Following Housen and Kuiken (2009) as well as Skehan (2009), this study operationalized speech fluency as a construct consisting of three different subdimensions: speed fluency, breakdown fluency, and repair fluency (see Appendix F for coding guidelines for fluency measures).

*Speed fluency*. Speech rate—one of the best predictors of fluency (Kormos, 2006)—was used by counting the number of syllables produced per minute of original speech including pause time, fillers, and dysfluencies (Mehnert, 1998; Sangarun, 2005; Yuan & Ellis, 2003). However,

for the pauses that were longer than 3 seconds, I counted them only as 3 seconds and eliminated the remainder as suggested by Riggenbach (1991). This coding technique is, according to Riggenbach, to obtain a more representative index for speech rate since some of the speakers were found to have long pauses of over 3 seconds in their monologues, which are not the common pause length of the non-native speaker (NNS)'s articulation rate. The length of all responses and pauses were measured by a sound recording and editing software called Audacity and rounded off to whole seconds.

*Breakdown fluency*. This component was measured by the number of pauses per minute (Elder & Iwashita, 2005; Kormos, 2006). All pauses greater than 1 second were counted (Foster & Skehan, 1996; Skehan & Foster, 1997)), and unfilled and filled pauses were calculated separately. Additionally, previous research found a difference in the positions of pauses between native and non-native speakers (Skehan, 2009; Skehan & Foster, 2005; Tavakoli, 2011), suggesting that the characterization of fluency needs to be deeper. As such, I also counted the number of unfilled pauses at the end and in the middle of clauses per minute to detect the pausing patterns across the experimental groups.

*Repair fluency*. This measure was operationalized as the number of dysfluencies such as reformulations or self-corrections, repetitions, replacements, and false starts (Foster & Skehan, 1996; Kormos, 2006; Skehan & Foster, 1999). Specifically, like counting the number of pauses per minute, the number of dysfluency phenomena was divided by total speaking time measured in seconds as in Elder and Iwashita's (2005) study and then multipled by 60 to obtain the number per minute.

Although the number of filled and unfilled pauses and other dysfluencies do not prove to be good indicators of perceived oral fluency (Kormos & Dénes, 2004; Lennon, 1990), they can

reflect a subdimension of fluency, and more importantly, they may signal the attention learners pay to linguistic form while attempting to monitor their production (Adams & Ross-Feldman, 2008). Thus, these measures were of great help to the current study whose purpose was to examine whether the participants could attend to what they were told to.

### *Complexity measures*

Complexity in this study was operationalized to include both syntactic complexity and lexical richness. I employed measures for overall complexity, phrasal complexity, subordination, and structural variety because they were all believed to represent distinct and complementary qualities of such a multidimensional construct as syntactic complexity (Norris & Ortega, 2009). I also measured lexical variety and lexical density as constituents of lexical richness (Laufer, 1991), which captures another aspect of complexity. These measures were computed after excluding all the fillers and dysfluencies from the total word count. Specifically, the measures of complexity can be detailed as follows (see Appendix G for further coding guidelines for complexity measures).

*Overall complexity*. Overall complexity was defined as mean length of T-unit or T-unit length, which was measured by dividing the number of words by the number of T-units in each narrative (Kawauchi, 2005; Mochizuki & Ortega, 2008). T-units rather than others were used partly because the task performance was monologic and contained few non-finite units (Foster, Tonkyn & Wigglesworth, 2000) and partly because they were commonly employed in analyzing writing, which was included in this study as a planning activity, thus easing the comparison between the written and spoken data.

*Phrasal complexity*. Phrasal complexity was defined as mean length of clause or clause length, which was measured by dividing the number of words by the number of clauses in each

narrative (Ishikawa, 2007). Notably, mean length of clause, which was previously employed as a measure of fluency rather than complexity, fundamentally differs from mean length of T-unit in that the former is not affected by the addition of subordinate clauses, but influenced by that of adjectives, adverbs, prepositional phrases, or non-finite verb phrases, thus it is currently proposed to serve as a specific measure that taps complexification at the phrasal level (Norris & Ortega, 2009).

*Subordination*. Amount of subordination was defined as the ratio of clauses to T-units, which was measured by dividing the number of clauses by the number of T-units in each narrative (Ellis & Yuan, 2005; Kawauchi, 2005; Sangarun, 2005). The minimum value for this score is thus 1.00 since every T-unit in this case contains only one clause.

*Structural variety*. Structural variety was defined as the variety of grammatical verb forms and was measured by counting the number of different verb forms used in each narrative. These forms included past tense aspects (e.g., simple past, past progressive, and past perfective), modality (e.g., should, should have, have to, had to), and voice (e.g., passive voice in the past) (Ellis & Yuan, 2005; Foster & Skehan, 1996). The variety of tensed forms is a valid measure for this study because it can gauge the use of such sophisticated forms as the past progressive and past perfective evidenced in the baseline data from 10 native speakers' narratives.

*Lexical variety*. Lexical variety was defined as the type-token ratio (TTR), which is the ratio in percent between the different lexemes in the text (e.g., nouns, pronouns, verbs, adjectives, articles, adverbs, prepositions, and conjunctions) and the total number of lexemes (Laufer, 1991; Laufer & Nation, 1995; Ortega, 1999; Robinson, 2001b). To measure this construct, the speech samples were entered on the computer in the form of lexemes. For instance, following Laufer (1991), I entered all inflected forms of verbs (e.g., *am, is, are, was, were,* 

*being*, and *been*) in the base form (e.g., *be*), treated homonyms as separate entries, entered an idiom (e.g., *put up with* or *rain cats and dogs*) as one item, and counted derivatives of base words as separate words (e.g., *happy, happily*, and *happiness*). Since the type-token ratio is strongly related to text length (Skehan, 2009), only the first 100 words (i.e., lexemes) of the narratives were considered for this measure.

*Lexical density*. Lexical density was defined as the percentage of lexical words in each narrative (i.e., nouns, verbs, adjectives, and adverbs) (Laufer, 1991; Mehnert, 1998). A narrative was considered "dense" when it contained many lexical words relative to the total number of words (i.e., lexemes).

#### Accuracy measures

Accuracy was operationalized by both general and specific measures including the number of errors per 100 words, the percentage of correct verb forms, target-like use of articles, target-like suppliance of plural -s, and the raw frequency of lexical errors in each narrative. These measures were computed after excluding all the fillers and dysfluencies from the total word count. Specifically, the measures of accuracy can be detailed as follows (see Appendix H for further coding guidelines for accuracy measures).

*Errors per 100 words*. To measure overall accuracy, I computed the number of errors per 100 words by dividing the number of errors by the number of words in each narrative and then multiplying the result by 100% as measured by Mehnert (1998) and Sangarun (2005). Consistently repeated errors due to wrong assumptions were counted only once as suggested by Mehnert (1998).

*Correct verb forms*. For a more specific measure of accuracy, I calculated the percentage of correct verb forms used in obligatory situations, following the measure employed by Wendel

(1997) and Yuan and Ellis (2003). These verb forms were first identified in relation to the use of tense, aspect, voice, modality, and subject-verb agreement. Then, the total number of correct verb forms was divided by the total number of verbs used in each narrative, and the result was multiplied by 100%.

*Target-like use (TLU) of articles*. Target-like use of articles was measured by dividing the number of accurately supplied articles by the number of obligatory contexts and inappropriately supplied articles, and then multiplying the result by 100% (Ortega, 1999; Pica, 1983; Robinson, 1995; Wigglesworth, 1997). By employing the TLU measure, suppliance in non-obligatory contexts (i.e., morpheme overgeneralization) can be taken into account, resulting in a reliable accuracy measure (Pica, 1983).

*Target-like use of the plural -s*. Target-like use of the plural –*s* was measured by dividing the number of accurately supplied plural -*s* by the number of obligatory contexts and inappropriately supplied plural -*s*, and then multiplying the result by 100% (Crookes, 1989; Pica, 1983; Wiggleworth, 1997). I also included the plural pronominal noun (e.g., *ones, others, the others*) because it forms its plural by use of the same –*s* suffix as regular plural nouns.

*Lexical errors*. I counted the raw number of lexical choice errors in each narrative (Foster & Skehan, 1996; Mehnert, 1998). It was assumed that planning time would facilitate the retrieval and activation of lexical items for the task performance, and thus resulting in fewer lexical choice errors (Mehnert, 1998). Lexical errors included serious deviations in pronunciation, meaning, grammatical form, word order, collocation, idioms, and expressions that interfered with the comprehensibility of the speech.

#### Analysis

# Analysis of pretests

For the grammar pretest, a test scoring machine was used to scan the answer sheets of 178 students I recruited. Among them, 126 students who correctly scored more than 40 out of 80 grammatical items were selected for the pronunciation pretest and task performance that followed a week later. For pronunciation check, my colleague--an experienced EFL teacher-- and I rated the participants' pronunciation recordings independently. We decided the pronunciation score of a participant based on a scale of 20 points for 20 correctly pronounced target words in the paragraph provided for reading aloud, and the inter-rater agreement for the pronunciation test reached 97%. Those who demonstrated poor pronunciation (i.e., achieving a score of 10 or less) were eliminated from the analysis. After screening the participants using the pretest scores to ensure that they were all at the equivalent level of proficiency, I had 97 participants left: 33 in the NP group, 32 in the P-W, and 32 in the P+W.

Then, I examined the task recordings and decided to exclude 7 more participants: 3 from the NP group, 2 from the P-W, and 2 from the P+W. These participants were eliminated because they had unintelligible recordings (i.e., they spoke too softly to understand), spoke too little (i.e., they performed the task for less than one minute instead of five minutes allotted), and did not plan as instructed (i.e., they kept silent instead of rehearsing, rehearsed instead of writing, and took notes instead of writing in full sentences). Finally, the data from a total of 90 participants, evenly distributed in three groups, were entered for analyses.

# Transcribing oral performances, rehearsals, and interviews

The participants' oral narratives and rehearsals were first transcribed verbatim and digitized by my colleague, who had received an MA's degree in Applied Linguistics at a U.S.

university and had been trained for collecting and coding data in this type of research before. Then, I double-checked all the transcribed data totaling more than 40,000 meaningful words, fillers, and dysfluencies. All the pauses, fillers, self-corrections, repetitions, replacements, and false starts were marked for easy count (see Appendix I for a speaking sample). As for the interviews, I first transcribed all the content in Vietnamese, then my colleague translated some selected excerpts into English.

### Analysis of oral performances

Approximately half of the participants performed the task within three to five minutes with an average of about three minutes per narrative (M = 172.82 seconds, SE = 6.26), so all the five minutes of each speech starting from the first word stated was chosen for analysis except those performances of less than one minute. In case the story was retold twice during performance, only the first time of retelling was counted. Moreover, whatever was added at the end of the story after a long pause of more than 3 seconds and unrelated or irrelevant ideas such as the speaker's comments and a lengthy life lesson drawn from the story were not counted either because they were not required in the task instructions. Following are two examples of cases that were excluded:

Example 1: "The story stop here. Can you understand this story? Can you infer any lesson from this story? Where after I tell story if you have any idea or if you have any misunderstand, please tell me. And I can have you to understand more clearly. Bye bye." Example 2: "So through this story we realize that do not hate everyone and especially do not hate dirty children because everyone want to go to school. They are not bad guy. That means if you hate someone, you will not get the good situation." To measure the fluency, complexity, and accuracy of the task performances, my colleague and I coded all the measures independently. Most of the measures for fluency, complexity, and accuracy were counted manually, using the basic coding schemes I developed following Polio's (1997, 2001) proposals (see Appendices F, G, and H for detailed coding guidelines). However, for the number of words and syllables, we first transcribed and entered the words as well as divided words into syllables manually, and then we used MS Word Count to count the words and syllables. Similarly, for the word types or different lexemes, we first identified all the lexemes independently and then extracted different lexemes with MS Word Duplicate Data Remove Software.

In addition to the detailed coding guidelines, I also developed several supplementary lists of more examples taken from the data to ease the coding. One of them was the list of more specific examples of what was counted as a T-unit or a clause. For instance, the two raters decided to count "*I think* ..." as a clause because there were some varieties like "*I don't think*", "*They think*", or "*They didn't think*", but did not count "*I mean*" because it sounds formulaic. There was another list of phrases entered as one lexeme. For example, we counted all the combinations like "*a lot of*", "*as soon as*", "*bus station*", "*get on*", and so on as one lexeme.

Further, it is necessary to emphasize that the decision on the errors and types of errors was carefully made together by two raters after consulting with several resources. For example, it was sometimes confusing for us--non-native researchers--to determine whether minor deviations from the rules of prescriptive grammar were acceptable in real communication or what phrase or expression was awkward (e.g., "*stand on a pavement*" versus "*stand on the pavement*", "*They waved their hand*" versus "*They waved their hands*", or "*enter into the bus*" versus "*enter the bus*". Thus, from the data collected, I also developed a list of all unacceptable use of language

and awkward phrases in consultation with four native speakers of English and Wordnik.com--an online dictionary that has collected a corpus of billions of words with example sentences. Some consistently repeated errors such as wrong article use (e.g., *"the bus thirty-three"* instead of *"bus thirty-three"*) or mispronunciations (e.g., *"pus"* instead of *"bus"*) were counted only once.

Inter-rater reliability was calculated, and the two raters reached an agreement of more than 90% for all measures (see Appendix J for a detailed table of inter-rater reliability). The scores were then entered into SPSS version 17.0.

A series of one-way ANOVAs was conducted to compare the differences between groups. For ANOVA to be reliable, all of its assumptions were tested. For instance, the Kolmogorov-Smirnov test was used to test the normal distribution of data, and it found that the distributions of most dependent variables for separate groups were normal. For non-normally distributed data such as the number of filled pauses per minute, the number of dysfluencies per minute, the number of verb forms per narrative, and the number of lexical errors per narrative, I used the Kruskal-Wallis test, a non-parametric counterpart of the one-way ANOVA, to test for differences between groups.

ANOVA is a reliable test for this study, for it is robust to violations of its assumptions (Field, 2009), especially given equal sample sizes. Post-hoc tests were also conducted to explore the relationships among groups and their effects on oral narratives. In this case, I used Tukey's test because of equal sample sizes and checked the findings with the Games-Howell procedure when equal variances were not assumed (Field, 2009). A significance threshold  $\alpha$  was set at .05 for all statistical tests. The effect size  $\omega$  was also calculated and will be reported along with the significance values.

# Analysis of planning data

For the P-W group, approximately half of the participants rehearsed their performance twice, and the other half did it for 3-4 times. All the successfully recorded 10-minute rehearsals of 23 participants were transcribed and entered for analysis. Seven participants' planning data were not analyzed partly because they were planned in Vietnamese or in mixed language and partly because they were improperly recorded (e.g., Four participants forgot to turn on the recording software at the beginning of the planning period until the researcher found out).

For the P+W group, the written narratives of 26 participants were digitized with underlines for insertions and strikethroughs for deletions to keep track of the changes the participant made (see Appendix I for a writing sample). Four participants wrote less than 100 words, thus their writings during planning were eliminated from analysis. While planning, the participants were told not to write in pencil because it would be hard to trace their corrections or changes. Approximately 4,000 written words of planning were entered for analysis.

The spoken as well as written data recorded during planning time were examined for instances of self-repair to investigate how often the planners chose to focus on form (FoF) or focus on meaning (FoM). For the rehearsals, self-repairs are defined to encompass all dysfluency phenomena such as reformulations/self-corrections, substitutive repetitions, replacements, and false starts because they may give us some clues about the form or meaning they notice. Replacements and false starts were counted as FoM, and reformulations or self-corrections were seen as FoF (see Appendix F for how to code replacements, false starts, and reformulations). However, in case of a false start which contained a reformulation, the false start and the reformulation were counted separately. According to Bygate (1996), a verbatim repetition occurs when hesitating to gain time to find an appropriate word, while a substitutive repetition takes

place when correcting a lexical or grammatical feature. However, exact repetitions were excluded from analysis in the present study as we could not tell whether the planner focused on form or meaning. For the writings, self-repairs were counted based on the insertions and deletions because they were clear evidence of FoF or FoM. All the pronunciation and spelling corrections were counted as FoF.

In general, instances of self-repair in the planning with rehearsals and writing were coded according to the target of the self-repair (e.g., focus on form or meaning), following Adams and Ross-Feldman (2008). Specifically, if the errors arise from a difficulty with formal linguistic aspects such as syntax, morphology, or pronunciation, the target of the self-repair is coded as Form. If learners are primarily engaged in how to express an intended meaning, the target of the self-repair is coded as Meaning. Then, proportion scores are calculated for each individual for self-repair (e.g., the number of FoF self-repairs is divided by the total number of self-repairs).

Inter-coder reliability was above .95 for all the counts of self-repairs in the planning data as FoF or FoM.

Finally, to have a more thorough understanding of the relationship between planning and performance, I further analyzed the oral and written planning data and the final oral narratives of four participants, one high achieving and one low achieving each from the P-W and P+W groups. The selection was based on the high or low scores of fluency, complexity, and accuracy measures of the participants' performance.

## Analysis of interview data

Interview responses from 10 participants of each group were coded independently by two coders according to the themes directed by the interview questions (see Appendix E) and any new theme emerging from the interview. Inter-coder reliability was high (r = .85, p < .01).

### CHAPTER 4: RESULTS

RQ 1: What effects does planning with and without writing have on the fluency of an L2 oral narrative?

# Speed fluency

As shown in Table 3, there is an increase in speed fluency measured by the number of syllables per minute, with the no-planning (NP) group being the least fluent (M = 92.64, SD = 18.44) and the P+W group the most fluent (M = 108.11, SD = 19.83). This suggests that both planning groups were likely to perform the task more fluently than the NP group.

Table 3: Descriptive statistics for speed fluency measure

		2			95% Confide	ence Interval
Measur	е	N	М	SD	Lower Bound	Upper Bound
Syllables/min.	NP	30	92.64	18.44	85.75	99.52
	P-W	30	100.05	16.64	93.84	106.27
	P+W	30	108.11	19.83	100.71	115.52
	Total	90	100.27	19.22	96.24	104.29

*Note.* NP = no-planning group, P-W = planning-without-writing group, and P+W = planning-with-writing group.

One-way ANOVA results revealed that there was an overall significant medium effect on the groups' ability to produce the number of syllables per minute, F(2, 87) = 5.34, p < .01,  $\omega =$ .30, (see Table 4). The results of Tukey's *post hoc* test further indicated that only the P+W group produced a significantly greater number of syllables per minute than the NP (p < .01), whereas there was no significant difference between the P-W group and the NP group (p = .266). The P-W group and the P+W group did not significantly differ from each other in this aspect (p = .211). In other words, those who planned their oral narratives by writing out the story in complete sentences spoke faster than those who were not allowed time to plan.

Measure	Source	SS	df	MS	F	ω	р
Syllables/min.	Between Groups	3594.03	2	1797.02	5.34	.30	.007
	Within Groups	29292.12	87	336.69			
	Total	32886.16	89				

Table 4: One-way ANOVA results for speed fluency measure

#### Breakdown fluency

An overall examination of the group means of the breakdown fluency measures in Table 5 shows that the planners tended to pause less than the non-planners for all of the measures. Interestingly, for the total pauses per minute, the P-W group paused least (M = 13.88, SD = 4.57) compared to the NP and P+W groups (M = 18.16, SD = 5.11 and M = 16.11, SD = 7.52, respectively). Likewise, for the number of filled pauses per minute, the P-W group used less fillers (M = 5.94; SD = 5.80) than the NP and P+W groups (M = 9.43, SD = 6.61 and M = 9.41, SD = 8.11, respectively). Another noteworthy observation was that the P+W group paused in the middle of clauses less (M = 3.30, SD = 2.05) than the NP and P-W groups (M = 5.11, SD = 2.05 and M = 4.51, SD = 2.77, respectively). However, for the number of unfilled pauses at the end of clauses, all the three group means for the NP (M = 3.62, SD = 1.56), P-W (M = 3.42, SD = 1.42), and P+W (M = 3.40, SD = 1.71) appeared to be roughly equal.

Measure					95% Confide	ence Interval
(per minute	e)	Ν	М	SD	Lower Bound	Upper Bound
Total Pauses	NP	30	18.16	5.11	16.26	20.07
	P-W	30	13.88	4.57	12.17	15.58
	P+W	30	16.11	7.52	13.31	18.92
	Total	90	16.05	6.07	14.78	17.32
Filled Pauses	NP	30	9.43	6.61	6.97	11.90
	P-W	30	5.94	5.80	3.77	8.11
	P+W	30	9.41	8.11	6.39	12.44
	Total	90	8.26	7.02	6.79	9.73
Unfilled Pauses	NP	30	8.73	2.83	7.67	9.79
	P-W	30	7.94	3.38	6.67	9.20
	P+W	30	6.70	3.02	5.57	7.83
	Total	90	7.79	3.16	7.13	8.45
Unfilled Pauses –	NP	30	3.62	1.56	3.04	4.20
End of Clause	P-W	30	3.42	1.42	2.89	3.95
	P+W	30	3.40	1.71	2.76	4.04
	Total	90	3.48	1.55	3.15	3.81
Unfilled Pauses –	NP	30	5.11	2.05	4.35	5.87
Mid-Clause	P-W	30	4.51	2.77	3.48	5.55
	P+W	30	3.30	2.05	2.54	4.07
	Total	90	4.31	2.41	3.80	4.81

Table 5: Descriptive statistics for breakdown fluency measures

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planningwith-writing group.

Table 6 represented the one-way ANOVA results of most breakdown fluency measures except for the measure on the number of filled pauses per minute because the data of this variable violated the assumption of normal distribution, and thus the Kruskal Wallis test was used instead. As in Table 6, there was an overall significant small-to-medium effect for planning on three breakdown fluency measures: the total number of pauses per minute, F(2, 87) = 4.00, p < .05,  $\omega$  = .25, the total number of unfilled pauses per minute, F(2, 87) = 3.29, p < .05,  $\omega = .22$ , and the number of unfilled pauses in the middle of clauses per minute, F(2, 87) = 4.77, p < .05,  $\omega = .28$ . For the number of unfilled pauses at the end of clauses per minute, the finding was not significant, F(2, 87) = .18, p = .836,  $\omega = .14$ .

Measure			-				-
(per minute)	Source	SS	df	MS	F	ω	p
Total Pauses	Between Groups	275.99	2	137.99	4.00	.25	.022
	Within Groups	3002.67	87	34.51			
	Total	3278.66	89				
Unfilled Pauses	Between Groups	62.70	2	31.35	3.29	.22	.042
	Within Groups	828.83	87	9.53			
	Total	891.53	89				
Unfilled Pauses –	Between Groups	.88	2	.44	.18	.14	.836
End of Clause	Within Groups	213.62	87	2.46			
	Total	214.51	89				
Unfilled Pauses –	Between Groups	50.95	50.95 2 25.48 4.77		4.77	.28	.011
Mid-Clause	Within Groups	465.17	87	5.35			
	Total	516.12	89				

Table 6: One-way ANOVA results for breakdown fluency measures

The results of Tukey's *post hoc* test in Table 7 further indicated that the P-W group generally has a significant smaller number of pauses per minute than the NP (p < .05), whereas the P+W group had significantly fewer unfilled pauses than the NP (p < .05), and this finding resulted from the fact that the P+W paused in the middle of clauses significantly less than the NP (p < .01).

Measure	Comparison	Mean		-	95% Confid	ence Interval
(per minute)	(I vs. J)	(I - J)	SE	р	Lower Bound	Upper Bound
Total Pauses*	P-W vs. NP	-4.29	1.52	.016	-7.91	67
	P+W vs. NP	-2.05	1.52	.371	-5.67	1.57
	P+W vs. P-W	2.24	1.52	.307	-1.38	5.86
Unfilled Pauses	P-W vs. NP	79	.80	.581	-2.69	1.11
	P+W vs. NP	-2.03	.80	.034	-3.93	13
	P+W vs. P-W	-1.23	.80	.273	-3.14	.67
Unfilled Pauses – Mid-Clause	P-W vs. NP	60	.60	.581	-2.02	.83
	P+W vs. NP	-1.81	.60	.009	-3.23	39
	P+W vs. P-W	-1.21	.60	.111	-2.64	.21

Table 7: Tukey HSD comparisons for breakdown fluency measures

*Note*. \* Total pauses included both filled and unfilled pauses.

For the number of filled pauses per minute, whose data were not normally distributed, it was noted that the mean rank was lower in the P-W group (35.90) than in the NP and P+W groups (51.32 and 49.28, respectively). The Kruskal-Wallis test results indicated that filled pauses were significantly affected by pre-task planning, H(2) = 6.168, p = .046. Mann-Whitney tests were used to follow up this finding with a Bonferroni correction for all effects at a .0167 (i.e., .05/3) level of significance. These pairwise comparisons showed that the P-W (Mdn = 3.79) had significantly fewer fillers per minute of their performance than the NP group (Mdn = 9.57), U = 299, p = .013 (one-tailed), r = -.29. However, for the measure of fillers, the P+W group (Mdn = 6.11) did not seem to differ from the NP group (Mdn = 9.57), U = 426.50, p = .364 (one-tailed), r = -.045. Again, there was no significant difference between the P-W and P+W groups in the number of fillers per minute of their performance, U = 313, p = .043 (two-tailed), r = -.26.

In brief, planning without writing had a significant effect on the total number of pauses per minute of speech, especially filled pauses, while planning with writing significantly affected the number of pauses in the middle of a clause.

### *Repair fluency*

Repair fluency was operationalized as four measures per minute: number of repetitions, reformulations, replacements, and false starts. In addition, the total number of repairs per minute including the four above-mentioned measures was calculated to obtain a general picture of repair fluency. An overall examination of the group means of the repair fluency measures in Table 8 shows that the P-W group tended to repair least and the NP repaired most. In fact, the P-W group made the smallest number of total repairs (M = 6.15, SD = 3.44) while the NP had the most repairs (M = 7.94, SD = 4.14). Likewise, for the number of repetitions per minute, the NP group made the most repetitions (M = 4.56; SD = 2.84), whereas the P-W repeated least (M = 3.11, SD = 2.17). Again, the P-W group made the least replacements (M = .60, SD = .47) while the NP group replaced the most lexical items (M = .67, SD = .68).

However, for false starts, the P+W group abandoned their utterances before completion (M = .37, SD = .46) more frequently than the other two groups, P-W (M = .41, SD = .48) and NP (M = .57, SD = .54). For the number of reformulations per minute, the P+W (M = 2.46, SD = 1.57) seemed to self correct their speech more than the NP (M = 2.15, SD = 1.28) and the P-W (M = 2.04, SD = 1.22).

Among the repair fluency measures, the number of reformulations per minute was the only measure that had normally-distributed data. However, one-way ANOVA results showed no significant difference between the groups, F(2, 87) = .756, p = .473,  $\omega = .07$ .

Measure	e	-	-	-	95% Confide	ence Interval
(per minu	te)	Ν	М	SD	Lower Bound	Upper Bound
Total Repairs	NP	30	7.94	4.14	6.39	9.48
	P-W	30	6.15	3.44	4.87	7.44
	P+W	30	7.50	4.47	5.83	9.17
	Total	90	7.20	4.07	6.35	8.05
Repetitions	NP	30	4.56	2.84	3.50	5.61
1	P-W	30	3.11	2.17	2.29	3.91
	P+W	30	4.05	3.26	2.83	5.27
	Total	90	3.90	2.82	3.31	4.49
Reformulations	NP	30	2.15	1.28	1.67	2.63
	P-W	30	2.04	1.22	1.58	2.49
	P+W	30	2.46	1.57	1.87	3.05
	Total	90	2.22	1.36	1.93	2.50
Replacements	NP	30	.67	.68	.41	.92
1	P-W	30	.60	.47	.42	.77
	P+W	30	.63	.62	.40	.86
	Total	90	.63	.59	.51	.75
False Starts	NP	30	.57	.54	.36	.77
	P-W	30	.41	.48	.23	.59
	P+W	30	.37	.46	.20	.53
	Total	90	.45	.49	.34	.55

 Table 8: Descriptive statistics for repair fluency measures

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planningwith-writing group.

For the measures of total repairs, repetitions, replacements, and false starts per minute, whose data were not normally distributed, Kruskal-Wallis tests were used to detect the differences between the groups. First, their mean ranks can be observed in Table 9. There appears to be a common pattern for the three groups on two measures, total repairs and repetitions, with the P-W group having the lowest mean ranks (39.37 and 38.57) and the NP group having the highest mean ranks (50.57 and 51.75). For replacements and false starts, there

was not such a clear pattern, but the P+W group had the lowest mean ranks (44.82 and 41.17) while the other two groups alternated having the highest mean ranks (46.15 for replacements in the P-W group and 51.22 for false starts in the NP group).

Table 9: Mean rankings for repair fluency measures

Measure	Ext	perimental Condition	on *
(per minute)	NP	P-W	P+W
Total Repairs	50.57	39.37	46.57
Repetitions	51.75	38.57	46.18
Replacements	45.53	46.15	44.82
False Starts	51.22	44.12	41.17
* 20			

\* n = 30 per group

 Table 10: Kruskal-Wallis test results for repair fluency measures

Measure (per minute)	Chi-Square	Df	р
Total Repairs	2.83	2	.24
Repetitions	3.85	2	.15
Replacements	.04	2	.98
False Starts	2.51	2	.29

Nevertheless, as indicated in Table 10, the Kruskal-Wallis tests revealed that there was no significant difference between groups for all these measures, total repairs (H(2) = 2.83, p = .24), repetitions (H(2) = 3.85, p = .15), replacements (H(2) = .04, p = .98), and false starts (H(2) = 2.51, p = .29). In short, pre-task planning seemed not to affect the repair fluency of the planning groups.

RQ 2: What effects does planning with and without writing have on the complexity of an L2 oral narrative?

					95% Confide	ence Interval
Measure		Ν	М	SD	Lower Bound	Upper Bound
Words / T-unit	NP	30	10.56	1.86	9.86	11.25
	P-W	30	11.81	2.23	10.98	12.65
	P+W	30	11.70	1.94	10.98	12.43
	Total	90	11.36	2.07	10.92	11.79
Words / Clause	NP	30	7.95	1.05	7.56	8.34
	P-W	30	8.77	1.16	8.33	9.21
	P+W	30	8.89	1.12	8.47	9.30
	Total	90	8.54	1.18	8.29	8.78
Clauses / T-unit	NP	30	1.33	.18	1.26	1.40
	P-W	30	1.35	.19	1.28	1.42
	P+W	30	1.32	.16	1.26	1.38
	Total	90	1.33	.17	1.30	1.37
Verb Forms	NP	30	2.17	1.23	1.71	2.63
	P-W	30	2.37	1.13	1.95	2.79
	P+W	30	2.87	1.31	2.38	3.35
	Total	90	2.47	1.25	2.21	2.73
Type-Token Ratio	NP	30	50.53	5.54	48.46	52.59
	P-W	30	53.10	3.81	51.68	54.52
	P+W	30	54.27	4.38	52.63	55.90
	Total	90	52.63	4.84	51.62	53.64
Content Words %	NP	30	47.82	4.41	46.18	49.47
	P-W	30	50.64	3.91	49.18	52.10
	P+W	30	50.15	4.23	48.57	51.73
	Total	90	49.54	4.32	48.63	50.44

Table 11: Descriptive statistics for complexity measures

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planning-

with-writing group.

A general examination of Table 11 showed that the P+W and P-W groups had higher means across the complexity measures than the NP group except the ratio of clauses to T-units, for which the groups means were approximately equal, and that the P+W and P-W groups alternated being the group that seemed to produce the most complex language.

One-way ANOVA results shown in Table 12 revealed that there was a significant smallto-medium effect for planning on four measures: words per T-unit, F(2, 87) = 3.56, p < .05,  $\omega =$ .23; words per clause, F(2, 87) = 6.34, p < .01,  $\omega = .33$ ; type-token ratio, F(2, 87) = 5.11, p < .01,  $\omega = .29$ ; and percentage of content words, F(2, 87) = 3.86, p < .05,  $\omega = .25$ . Tukey HSD comparisons in Table 13 located more specific differences between the groups, which are presented as follows.

Measure	Source	SS	df	MS	F	ω	р
Words / T-unit	Between Groups	28.96	2	14.48	3.56	.23	.033
	Within Groups	354.07	87	4.07			
	Total	383.03	89				
Words / Clause	Between Groups	15.64	2	7.82	6.34	.33	.003
	Within Groups	107.29	87	1.23			
	Total	122.93	89				
Clauses / T-unit	Between Groups	.01	2	.01	.21	.13	.810
	Within Groups	2.70	87	.03			
	Total	2.71	89				
Type-Token Ratio	Between Groups	219.57	2	109.79	5.11	.29	.008
	Within Groups	1867.96	87	21.47			
	Total	2087.54	89				
Content Words %	Between Groups	135.68	2	67.84	3.86	.25	.025
	Within Groups	1527.36	87	17.56			
	Total	1663.04	89				

Table 12: One-way ANOVA results for complexity measures

Maagura	Comparison	Mean		-	95% Confid	ence Interval
Weasure	(I vs. J)	(I - J)	SE	р	Lower Bound	Upper Bound
Words / T-unit	P-W vs. NP	1.26	.52	.047	.01	2.50
	P+W vs. NP	1.14	.52	.078	10	2.39
	P+W vs. P-W	11	.52	.974	-1.36	1.13
Words / Clause	P-W vs. NP	.82	.29	.014	.14	1.51
	P+W vs. NP	.94	.29	.004	.25	1.62
	P+W vs. P-W	.11	.29	.915	57	.80
Type-Token Ratio	P-W vs. NP	2.57	1.20	.086	28	5.43
	P+W vs. NP	3.74	1.20	.007	.89	6.59
	P+W vs. P-W	1.16	1.20	.595	-1.69	4.02
Content Words %	P-W vs. NP	2.81	1.08	.029	.23	5.39
	P+W vs. NP	2.33	1.08	.086	25	4.91
	P+W vs. P-W	49	1.08	.894	-3.07	2.09

Table 13: Tukey HSD comparisons for complexity measures

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planningwith-writing group.

For overall complexity, the P-W (M = 11.81, SD = 2.23) produced significantly more words per T-unit than the NP (M = 10.56, SD = 1.86) (p < .05), whereas the difference between the P+W (M = 11.70, SD = 1.94) and the NP was just close to significance (p = .078). The P+W and the P-W did not significantly differ in the number of words per T-unit in their speech (p = .974).

For phrasal complexity, both the P+W (M = 8.89, SD = 1.12) and the P-W (M = 8.77, SD = 1.16) produced significantly more words per clause than the NP (M = 7.95, SD = 1.05) (p < .01 and p < .05, respectively), whereas the two planning groups did not differ from each other significantly (p = .915).

For lexical variety, the P+W (M = 54.27, SD = 4.38) produced significantly more types (different lexemes) for the first 100 words than the NP (M = 50.53, SD = 5.54) (p < .01), whereas the difference between the P-W (M = 53.10, SD = 3.81) and the NP was just close to significance (p = .086). The P+W and the P-W did not significantly differ in the number of different lexemes per 100 words (p = .595).

For lexical density, the P-W (M = 50.64, SD = 3.91) produced significantly more lexical words per pruned narrative than the NP (M = 47.82, SD = 4.41) (p < .05), whereas the difference between the P+W (M = 50.15, SD = 4.23) and the NP was just close to significance (p = .086). Again, the P+W and the P-W did not significantly differ in the number of lexical words per narrative (p = .894).

For the number of different verb forms, both the P+W and P-W groups appeared to have more structural variety in their speech (M = 2.87, SD = 1.31 and M = 2.37, SD = 1.13, respectively) than the NP group (M = 2.17, SD = 1.23). Since the data of verb forms were not normally distributed, the Kruskal-Wallis test was used to detect the between-group differences. Like the group means, the mean ranks for the P+W group (52.78) and the P-W (43.68) were higher than that for the NP group (40.03). Again, it seemed that both planning groups produced more verb forms than the NP, but the Kruskal-Wallis test results indicated that there was no significant difference between groups (H(2) = 4.02, p = .134), suggesting that the three groups had equivalent structural variety in their performance.

In brief, this study gives evidence for a small-to-medium effect for pre-task planning on the complexity of L2 oral narratives. Planning with and without writing had a significant effect on phrasal complexity measured by the number of words per clause. However, a type of planning may be more sensitive to some measures than another, which was evidenced by the fact that planning without writing had significant effects on overall complexity measured by the number of words per T-unit and lexical variety measured by the percentage of content words, whereas planning with writing significantly affected lexical variety measured by the type-token ratio.

RQ 3: What effects does planning with and without writing have on the accuracy of an L2 oral narrative?

Table 14 presented a summary of descriptive statistics for the accuracy measures including the number of errors per 100 words, the percentage of correct verb forms, the target-like use of articles, and the target-like use of the plural -s, and lexical errors.

For overall accuracy measured by error per 100 words, a decrease in the number of errors could be observed in the narratives of the two planning groups (M = 10.75, SD = 4.12 for the P-W and M = 10.88, SD = 4.27 for the P+W) compared to the NP (M = 12.67, SD = 3.89). It was also noted that there was an increase in the percentage of correct verb forms in the speech of the planners (M = 74.29, SD = 11.36 for the P-W and M = 74.32, SD = 10.42 for the P+W) compared to the NP (M = 64.95, SD = 12.94). Also, both planning groups, the P+W and the P-W, made fewer lexical errors (M = 2.97, SD = 2.36 and M = 3.83, SD = 2.78, respectively) than the NP (M = 4.30, SD = 2.62). However, such a pattern could not be detected regarding the two specific measures of accuracy: the target-like use of articles and plural –*s*. For article use, the NP (M = 84.98, SD = 6.76) and even use more correct articles than the P+W (M = 80.54, SD = 15.54). For the plural –*s*, the NP (M = 67.48, SD = 23.65) appeared to have better target-like use of this morpheme than the P-W (M = 65.11, SD = 30.52) and the P+W (M = 62.33, SD = 25.88).

		_	-	-	95% Confidence Interva	
Measure		N	M	SD	Lower Bound	Upper Bound
Errors / 100 words	NP	30	12.67	3.89	11.22	14.12
	P-W	30	10.75	4.12	9.22	12.29
	P+W	30	10.88	4.27	9.28	12.47
	Total	90	11.43	4.14	10.57	12.30
Correct Verb Form	NP	30	64.95	12.94	60.12	69.79
	P-W	30	74.29	11.36	70.04	78.53
	P+W	30	74.32	10.42	70.43	78.21
	Total	90	71.19	12.31	68.61	73.77
Correct Article Use	NP	30	84.40	8.70	81.15	87.64
	P-W	30	84.98	6.76	82.45	87.50
	P+W	30	80.54	15.54	74.74	86.34
	Total	90	83.30	11.05	80.99	85.62
Correct Plural Noun	NP	30	67.48	23.65	58.65	76.31
	P-W	30	65.11	30.52	53.71	76.51
	P+W	30	62.33	25.88	52.66	71.99
	Total	90	64.97	26.62	59.40	70.55
Lexical Errors	NP	30	4.30	2.62	3.32	5.28
	P-W	30	3.83	2.78	2.80	4.87
	P+W	30	2.97	2.36	2.09	3.85
	Total	90	3.70	2.62	3.15	4.25

 Table 14: Descriptive statistics for accuracy measures

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planningwith-writing group.

One-way ANOVA results shown in Table 15 revealed that there was a significant medium effect for planning on the percentage of verb forms, F(2, 87) = 6.48, p < .01,  $\omega = .33$ , whereas no significant difference between groups was found in the number or errors per 100 words, F(2, 87) = 2.05, p = .135,  $\omega = .15$ , the target-like use of articles, F(2, 87) = 1.44, p = .242,  $\omega = .10$ , and the target-like use of the plural -s, F(2, 87) = .28, p = .759,  $\omega = .13$ .

Measure	Source	SS	df	MS	F	ω	p
Errors / 100 words	Between Groups	68.82	2	34.41	2.05	.15	.135
	Within Groups	1459.17	87	16.77			
	Total	1527.99	89				
Correct Verb Form	Between Groups	1748.54	2	874.27	6.48	.33	.002
	Within Groups	11744.79	87	135.00			
	Total	13493.33	89				
Correct Article Use	Between Groups	348.84	2	174.42	1.44	.10	.242
	Within Groups	10522.32	87	120.95			
	Total	10871.17	89				
Correct Plural Noun	Between Groups	398.63	2	199.32	.28	.13	.759
	Within Groups	62661.51	87	720.25			
	Total	63060.14	89				

Table 15: One-way ANOVA results for accuracy measures

Tukey HSD comparisons in Table 16 located an interesting significant difference in verb form percentage between the two planning groups and the control group (p < .01), indicating that the planners used significantly more correct verb forms than the non-planners.

Table 16: Tukey HSD comparisons for accuracy measures

Measure	Mean Comparison Difference				95% Confidence Interval		
	(I vs. J)	(I - J)	SE	р	Lower Bound	Upper Bound	
Correct Verb Form	P-W vs. NP	9.33	3.00	.007	2.18	16.49	
	P+W vs. NP	9.37	3.00	.007	2.21	16.52	
	P+W vs. P-W	.04	3.00	1.000	-7.12	7.19	

*Note*. NP = no-planning group, P-W = planning-without-writing group, and P+W = planningwith-writing group. For the number of lexical errors per pruned narrative, whose data were not normally distributed, it was noted that the mean ranks decreased from the NP group (53.65), then the P-W (46.20) to the P+W (36.65). The Kruskal-Wallis test results indicated that lexical errors were significantly affected by pre-task planning, H(2) = 6.573, p = .037. Mann-Whitney tests were used to follow up this finding with a Bonferroni correction for all effects at a .0167 (i.e., .05/3) level of significance. These pairwise comparisons showed that the P-W group (Mdn = 3.00) did not seem to differ in making lexical errors from the NP (Mdn = 4.00), U = 381.00, p = .295, r = .14, or from the P+W (Mdn = 2.00), U = 360.00, p = .179, r = .17. However, a significant difference was found between the P+W and the NP group, U = 274.50, p = .008, r = ..34, meaning that the P+W group significantly made fewer lexical errors in their performance than the NP, while the P-W did not.

To sum up, pre-task planning with and without writing in this study had a significant effect on the accuracy of oral narratives measured by the number of correct verb forms. Especially, planning with writing significantly reduced the number of lexical errors in the planners' oral narratives.

#### RQ 4: What do learners attend to during planning?

There are three types of data that provide answers to the fourth research question. First, the results from an analysis of self-repairs during planning with and without writing will be reported separately and then compared to see if there is any common pattern of attention in the way the participants planned their speech. For the interview data, themes will be presented following the interview questions to gain more insights into what the participants chose to attend to and what they did not. Further, for a more detailed understanding of how the participants'

narratives changed from planning to performance, I will describe the oral and written planning data and the final oral narratives of four special participants, one high achieving and one low achieving each from the P-W and P+W groups.

# Results from analysis of self-repairs in the planning data

For planning without writing (P-W), almost all the planners (n = 23) used 5-10 minutes for rehearsing (M = 7.73, SD = 1.89), with only one exception of planning for nearly 3 minutes. On the average, a participant in this group spoke approximately 500 words (excluding fillers and verbatim repetitions) in his/her rehearsal (M = 492.83, SD = 186.67). Nearly half of the planners rehearsed their performance twice, and the other half did it for 3-4 times. Only three speakers had one rehearsal cycle. Most of the planners rehearsed their speech in English; however, one participant planned it completely in Vietnamese; another first did it in Vietnamese and then in English; and another had three rehearsal cycles with the first time in Vietnamese, the second one in mixed language, and the third one in complete English.

Generally, the percentage of focus on meaning (FoM) (i.e., the ratio of FoM to total self-repairs) in the rehearsal (M = .53, SD = .16) was slightly higher than that of focus on form (FoF) (i.e., the ratio of FoF to total self-repairs) (M = .47, SD = .16), suggesting that the P-W planners spent more time focusing on meaning than focusing on form. Further examination of categories of focus based on the coding guidelines for classifying errors (see Appendix H) showed that focus on lexical items accounted for 44% of total self-repairs (M = .44, SD = .15), which was the highest percentage compared to that of some other categories such as verb forms (19%), articles (5%), plural nouns (3%), and prepositions (3%). It means that vocabulary was the top priority to the P-W planners.
For planning with writing (P+W), all the planners (n = 26) used up 10 minutes of planning and wrote more than 100 words (excluding self-repairs) in their narratives. Half of the planners could write more than 130 words per pruned narrative (M = 133.69, SD = 19.41). On the average, they wrote nearly 12 words per T-unit (M = 11.87, SD = 2.41).

Interestingly, unlike the P-W group, the P+W planners tended to focus on form (M = .55, SD = .20) more than focus on meaning (M = .45, SD = .20) in their written narratives. Although this was not a big difference, it might indicate that writing enables learners to focus on form better than rehearsing. Moreover, a closer examination of categories of focus in planning with writing revealed a pattern similar to that of the planning with rehearsals. It means that the P+W planners focused most on lexical items (accounting for 37% of total self-repairs), then verb forms (22%), articles (11%), prepositions (6%), and plural nouns (1%). It also seems that the P+W planners allocated more attention to grammatical forms such as verb forms, articles, and prepositions than their P-W counterparts.

### Results from qualitative analysis of the interview data

For the interview data, I will basically report the results by themes categorized under each of the questions regarding the planners' allocation of attention, concerns, and difficulties during planning. I will also present their perceived success in task performance, ability to use what was planned, and favorite way of planning. Due to some confusion with the question about the difficulties of the type of planning experienced, many participants reported their difficulty with the task; therefore, I will present their responses to this question together with their biggest concerns during planning. All the interview questions were designed for the planners in the P-W and P+W groups; however, I also had ten speakers from the NP group volunteer to answer the

questions about the extent of success in their oral performance and their favorite way of planning.

Interview question 1: "What were you thinking about while planning?"

An overall observation of the interview data indicated that there was a similar pattern of attention in two groups of planning – planning without writing (P-W) and planning with writing (P+W). Most planners appeared to devote more time to planning the content and vocabulary than planning grammar and pronunciation. Nine out of ten planners (90%) in each planning group considered ideas and vocabulary their top priorities, while only 60% P-W planners and 70% P+W ones answered that they attended to grammar during planning. Five out of ten P-W planners (50%) said that they paid attention to pronunciation. Since the P+W group planned with writing, pronunciation was not the linguistic aspect this group attended to. In general, most participants thought about the content, vocabulary, and grammar while planning and tried to plan their oral performance following the instructions as reflected in the following examples:

I spent most time looking for the details and vocabulary to express the ideas. Once I had the right words, I then thought of the appropriate structures. [Participant 0303]<sup>1</sup> I paid most attention to content and vocabulary because we were required to tell a story in a way to make it understood by a listener who had never heard the story before. [Participant 0209]

Further examination of the planners' revelation details major similarities and minor differences in the way they processed the four aspects: content, vocabulary, grammar, and pronunciation.

<sup>&</sup>lt;sup>1</sup> Each participant was given a 4-digit code with the first two digits denoting his/her planning group and the last two digits denoting his/her identifying number in the group.

Attention to content. Unlike the NP group, given limited planning time of 30 seconds,

who "*only chose to focus on main points and ignore details*" [Participant 0102], the planners of the other two groups had more time to observe the pictures and thus could explore more details, especially the important differences between the two groups of boys, two bus numbers, and two clocks:

I first looked at each picture, attended to different colors and the order of the pictures. I tried to understand the details. Then, I practiced telling the story picture by picture. For the first time of telling, I told some details; for the second time, I added more details and compared the differences between pictures. At first, I saw only one bus – bus 26, but when I told the story the second time, I saw bus 33. [Participant 0206] I paid attention to the different colors to distinguish two groups of boys. I focused on three small boys and four young men because the small boys were the main characters

and the four men were the cause of the problem. I wondered where the children went and what they did. [Participant 0304]

The planners even used their rich imagination to make up new details for the story to sound interesting to the listener as revealed by one planner:

I wondered why they did not get on bus 26, but got on bus 33. They may have foreseen something wrong with bus 26 with their sensitivity... I did not know where the three boys would go, so I said that they were travelling to the countryside. [Participant 0304]

While the non-planners did not have enough time to notice the connections between the pictures and rushed to describe whatever they saw in each picture separately, the planners could organize the ideas during planning to retell a coherent story. One planner even attempted to express the feeling of the three small boys in red shirts and thought of the meaning of the story:

I related it to the proverb "More haste, less speed". [Participant 0310]

A few planners had to rely on their L1 to plan the message and search for L2 vocabulary and structures to convey it as evident in this example:

I first told the story in Vietnamese as my habit, then looked for the words in English, and where to put an article. If there had been time left, I would have tried to rehearse in English as much as possible. [Participant 0207]

*Attention to vocabulary*. Planners spent most time searching for the contextually appropriate words to express ideas because they thought that vocabulary was the most important and it helped them speak more easily. Many of them complained that they were unable to recall some simple words like "*mud*", "*push*", "*splash water*", "*bus conductor*", "*the winding road*", and so on. The planners, especially the P+W group, were aware of varying their word choice to make the story interesting. Two planners noted:

I tried to find the right words for the details in the pictures because it was important to tell the story will all the details ... Since this is an interesting story, I would use adjectives to describe the characteristics of these students and the people sitting on the bus as well as the bus driver ... [Participant 0209]

I paused long to search for proper vocabulary to express ideas and think of synonyms or antonyms to make it [the story] less boring. For example, small boys could be expressed by little boys; however, for 'get on', I did not know what to replace. [Participant 0301]

*Attention to grammar.* Planners did attend to grammar, mostly verb tenses, and they thought that the past tense would be more appropriate for telling a story. The other grammatical points they mentioned in the interview included structures, the definite article, singular/plural

nouns, and singular third-person subject and verb agreement. The following responses explain why planners focused on the past tense and sentence structures:

First, I thought that I would use the present tense, but later thought that I should use the past tense when telling a story because of being guided with "One day ..." to start the story, so I corrected the mistakes. [Participant 0208]

*I attended to sentence structures because there was a need to pause at appropriate points to make the sentences easy to understand.* [Participant 0203]

The reasons for the planners' attention to grammar varied from the task demands to personal consciousness. Although some students did not understand why they attended to tenses why planning, some others admitted that they were not good at grammar use and often made mistakes with tenses, so they attended to this grammatical point as a habit. However, the most interesting reason is that grammatical accuracy was emphasized by high school teachers, and students gradually established a habit of focusing on form while writing and speaking as reflected by many participants. Some students showed to be strongly aware of grammatical accuracy in speaking as honestly revealed in this response:

... because in the past when I was in high school, every time I said something wrong, I felt embarrassed and couldn't continue. I learned a lot of grammar then, but did not learn how to speak English. Every time I used wrong grammar, the teacher corrected .... I was born and studied in a small town. My teacher, who was claimed to be the best in high school, was not a good speaker of English, so he taught grammar for most of the time. [Participant 0208]

Attention to pronunciation. Fifty percent of those who rehearsed the oral performance within 10 minutes of planning showed their concerns about pronunciation. Two of the

participants admitted that they tended to focus on pronunciation as a habit. They were afraid of mispronunciation because it would lead to misunderstanding or no understanding, whereas they had pressure of making the story as interesting as possible. Besides, Vietnamese learners of English tend to drop the final consonants in English pronunciation, thus several P-W planners were often aware of some inflectional suffixes. The following excerpts illustrate such reasons:

Most of my attention was given to pronunciation because if I use wrong grammar or make mistakes in using structures, people can still understand me, but if I have wrong pronunciation, there will be no understanding. [Participant 0201]

I paid attention to plural noun endings because we often make these mistakes and because while learning in the speaking/listening courses, I noticed that my classmates often dropped the final consonants, and teachers often reminded them. Thus, I am aware of improving my pronunciation in this area. [Participant 0207]

Another reason for paying attention to pronunciation was that the planners cared about the listener. This concern was expressed as follows:

I cared about intonation because when you tell a story, if people do not like it, it is a failure. [Participant 0204]

Telling a story just for understanding is not difficult, but if required to make it interesting, I think my telling voice was not pleasant to listen to. If there had been a listener here, he must have fallen asleep. [Participant 0208]

For grammar and pronunciation, a few planners also revealed that they focused on these two aspects because they were their strengths or their weaknesses. For instance, one planner said:

*I was most concerned about grammar because I am not good at it.* [Participant 0304]

Interview question 2: "What were you most concerned about during planning?"

In general, more than half of the interviewees (60%) admitted that vocabulary was the biggest concern while planning, while only about a quarter (26%) considered exploring ideas to be the most difficult part, and the rest thought that grammar and pronunciation were the hardest, 12% and 2% respectively. The reasons for claiming vocabulary to be the biggest concern are clearly explained in these examples:

There were many points where it took so much time to think of vocabulary. Although I tried to find substitutes, but expressing an idea using a substitute word is not as good as an accurate word. [Participant 0304]

While planning, I could not get the word from memory right away or did not have the right word available. It was perhaps because I did not practice often. To take an example, for the word 'mud', I had to think hard for a long time, and then remembered. [Participant 0202]

It was obvious that the planners had to struggle with vocabulary search, but what they were most concerned about was sometimes not perceived as the biggest difficulty. For them, problems with finding the right word could be solved by using substitutes or longer expressions, and the biggest difficulty seemed to be what they could not achieve. Such a perception was shared in this remark:

The biggest difficulty is tenses and pronunciation of plural noun endings. Vocabulary is not the biggest difficulty because we can express one idea in another way; however, searching for vocabulary took most time. [Participant 0307]

Interview question 3: "Do you think you performed the oral task successfully?"

The NP group was the group that reported the most problems in their performance and admitted little success. Six out of ten interviewees (60%) replied that they did not have a satisfactory performance, and only one interviewee responded that she completed 60% of the story. Common difficulties that led to the non-planners' failure consisted of poor knowledge of vocabulary, misunderstanding the details, incoherent storytelling, wrong tense use, and so on. Two interviewees elaborated why vocabulary was the reason for their unsuccessful performance:

I was unable to express many details due to limited vocabulary. I just said several things in general ... When looking at each picture, I just described what I could see in it, and moved quickly to the next one. Finally, I found it wrong because it was in contrast with the beginning. [Participant 0109]

If given a topic, I think it would be easier to think about what to say in our own words, but telling a story like this, we had to say the required ideas, and I forgot the words to express the ideas in the pictures ... The pictures showed very common activities, but I could not remember the vocabulary. It was not because I did not know. [Participant 0105]

The non-planners were aware of using the past tense to retell a story, but because they had to struggle with understanding the pictures and completing the story, they failed to use the correct tense. This was one of the reasons for their dissatisfaction with their storytelling. One interviewee remarked:

When telling a story, we should use the simple past. I did not use the simple past through the story because the present tense is more simple and does not require verb conjugation. [Participant 0103] Without planning time, the non-planners could not understand the plot of the story and the contrast between the beginning and the ending of the story just by glancing at the pictures and they found them confusing, especially picture 5 and 6, as evidenced in the following examples:

Picture 5 is confusing. For example, the buses in the other pictures have their number plates, but not in picture 5, so I just had a glance at this picture and then found the numbers in picture 6. In pictures 3 and 4, I did not attend to the bus numbers; I only paid attention to three small boys in red shirts. In the last picture, I saw two buses but did not recognize two groups of boys. [Participant 0107]

I did not understand the story clearly, feeling so confused with pictures 5 and 6. At first glance, I thought about the unluckiness of the three boys on the street, such as being splashed with water, failing to get on the bus, and their bus broken on the way, but after looking at picture 6, I saw the three small boys laughing, then I recognized that the broken bus was the first one that they were not allowed to get on. [Participant 0108]

As for the P-W group, interviewees felt fairly satisfied with their performance. No interviewees reported having difficulty understanding the storyline. Two participants thought that they were quite successful in telling the story. Four others were not very pleased with their speaking because of some common problems such as inability to use as many words and ideas as planned, repeated structures and mispronunciations. Only one student admitted that he was not successful in telling the story because he could not make the story funny as it was. Some problems the P-W group encountered can be seen in the two examples that follow:

Within 5 minutes of speaking, I forgot many ideas planned for 10 minutes before. I was unable to speak much. There were many other ideas, but I did not have enough

vocabulary to express. I did not vary the structures, and repeated the same structures, making it boring. [Participant 0203]

The tense that I used was not appropriate; I should have used the past tense because the story happened, but sometimes I used the simple present and thought it wrong, so I felt ashamed and made more mistakes When making mistakes in pronunciation, I paused and corrected, thus wasting time. [Participant 0206]

Finally, among the planners of the P+W group, five out of ten interviewees (50%) believed that they did not succeed in telling the story. A common reason for this perception of failure was that they did not have the right words to express all the details, thus leading to many repetitions and hesitations. Another typical complaint of this group was that they could not remember some details despite trying to remember what had been written. One student also admitted that she told a lengthy and choppy story and the listener might find it boring. She added:

I should have used more adjectives for the listener to visualize the characters' mood. I did not have an expressive voice either. [Participant 0305]

Some other reasons included worries about tenses and story perfection as in the following remark:

When speaking, I couldn't tell the significance of the story. Telling a story requires accurate use of past tenses, but I couldn't. [Participant 0301]

Only one student was confident that he successfully completed the storytelling.

Task unfamiliarity appeared to be another cause for the participants' perception of lack of success. Five interviewees from all the three groups revealed that they had never told a story

based on a series of pictures before; as a result, they were very confused with the task at the beginning. One interviewee from the NP group commented:

This is the first time for the past four years I have had a chance to tell a story based on pictures, so I could not do it well. A quick look at the pictures did not give me enough time to understand them. First, I thought I understood, but when I began to speak, I found out there were many points I did not understand and wanted to skip these points, so I had only a few ideas left. [Participant 0107]

Although the extent of the interviewees' perceived success was definitely subjective, their revelation of the reasons showed that the type of planning had some influence on their performance. The planners seemed more positive about their speaking than the non-planners and the P-W group appeared to be the most successful in completing the task. It is evidenced that the planners had the planning time to access their explicit knowledge to select and refine their vocabulary, structures, and pronunciation. While the non-planners considered story completion to be their priority to the exclusion of grammatical accuracy, the planners tended to perfect their work by varying word choice and structures, and even trying to use an interesting tone to attract the listeners' attention.

Interview question 4: "Could you use what you had planned in your narrative?"

Six out of ten interviewees from the P-W group reported that they were able to speak about 70-90% of the sentences prepared. They could avoid many mispronunciations and even used intonation to make the storytelling more interesting. However, some regretted missing some details when speaking and making a few changes in word choice, as reflected in the following comments:

I forgot the detail that the conductor said the bus was full and told them to wait for another bus ... I was afraid of running out of time, so I spoke fast, and when speaking fast, my performance was not like what had been planned before. For instance, I used different words, while structures remained the same. [Participant 0204] I could not use all the words planned possibly because of losing confidence or feeling lost. It happened when I stopped for about 10 seconds to look for more ideas to keep telling. [Participant 0209]

Similar to the P-W group, 60% interviewees from the P+W group responded that they could use what they had planned. Many admitted trying to remember their writing, and thus being able to use all the words written out. They could also self-correct the grammatical mistakes, especially the past tense and articles. More importantly, they seemed confident that they understood the main ideas and told the story in the right order. They could point out the difference between two buses and the contrast between the beginning and the end of the story, which most non-planners could not do, as one interviewee reflected:

While speaking, I tried to remember what I had planned in writing. I also paid attention to the colored features in the pictures to find out more details to say. For example, I saw two buses and realized the ending in contrast with the beginning. [Participant 0304]

Nonetheless, one student confessed that his speech was not a great success because while writing he had many ideas, but when speaking he spoke differently from what prepared especially for the last pictures and was unable to express all thoughts, and his oral performance was worse than his writing:

I was able to speak ok for the first half of the story because I remembered the writing, then got stuck for the second half, so I spoke differently from the writing, but in general there was not much difference in terms of ideas. [Participant 0303]

Interview question 5: "What are the benefits of the type of planning you experienced?"

For the P-W group, many interviewees reported that planning with rehearsal helped them speak more fluently because they could remember the main points, vocabulary, tenses, and their anxiety was relieved. One interviewee expressed the benefit of this type of planning as follows:

Without planning, it would be hard to speak smoothly. I could remember some main points. I was less nervous. I rehearsed several times for 10 minutes, so I could remember the vocabulary and tenses. [Participant 0203]

Nevertheless, not all the planners could remember and use what they planned successfully, as one planner said:

For ten minutes of planning, I planned in Vietnamese and could speak more ideas, but within 5 minutes of speaking, I forgot many ideas planned 10 minutes before, so I couldn't express all. [Participant 0204]

For the P+W group, according to the interviewees, planning with writing encouraged coherent and fluent speaking, lexical and structural variety, and focus on form. First, given time to write before speaking, the planners could speak more fluently because they could remember the ideas. For example, a planner said:

Planning with writing helped with memory and thus helped speak more fluently. When writing, we organized the ideas, and we only orally repeated what we wrote because we could remember it. Without writing, we still have ideas to say, but not in a coherent way and with less confidence. For example, when speaking, if we expressed the ideas in a mess and repeatedly revise the order, we would lose confidence and could not keep going. [Participant 0306]

Second, writing was claimed to provide the planners time to select and vary vocabulary and structures as shown in one interviewee's response:

Writing is easier than speaking because we can wait to think about the appropriate words, and can adjust or vary grammatical structures, but when speaking we have to think fast, thus making more mistakes. If we are nervous, our utterances will be sporadic with no connections between ideas, but while writing, you can think more and add where necessary. [Participant 0304]

Third, some planners found it easy to see the mistakes when writing out the story on paper, whereas it is hard to do so when speaking. It means that writing can facilitate focus on form. One planner remarked:

I found it easy to see the mistakes when writing it out, but it might not be so when speaking ... I used all effort to complete the story with grammatical accuracy, but when speaking, we don't have to do so. Normally, when speaking, I only use the simple present, but when writing this story, I attended more to –ed endings and final consonants. [Participant 0302]

However, despite acknowledging the benefit of writing before speaking, a planner noted a disadvantage of planning with writing in comparison with speaking spontaneously.

Usually, we speak spontaneously without planning time, so we have no fluency and miss many details, and as such planning helps better speaking. If we have a chance to write out once what we are going to say, we may roughly remember ideas and speak faster after that. We often tend to remember what we write if we have a good memory. For this task, I could remember all my writing and when speaking. I even recognized and corrected some points that I did not write well. However, when speaking without planning by writing, we speak whatever we have. That is real speaking skill. If we write something, then we can remember and read it from memory, so it does not reflect the real ability to speak in a spontaneous situation. [Participant 0304]

Those who did not have a good memory seemed not to favor planning with writing as reflected in this remark:

If allowed 10 minutes just to look at the pictures and plan what to say in mind, it would be better than writing it out on paper because planning in mind and speaking out right away would be easier to remember. When writing out the story, I tried to remember it, and then when speaking I found out that I could not remember everything because the writing was taken away. While writing, I had many ideas, but then I spoke differently from what I prepared. It was worse than the writing because I could not express all thoughts. [Participant 0308]

Besides, the planners of the P+W group complained that writing was time-consuming, while they did not have enough time to tell all the details. As a result, they tried to compensate by adding more ideas in their oral performance.

Since I wrote at a rather slow pace and often paused long to search for vocabulary, 10 minutes of planning was not enough for me. When I could not finish writing up the last part of the story, I felt nervous. I was also afraid of forgetting what was written when speaking. The first half of my speaking was OK because I tried to remember. Planning by writing takes a lot of time. [Participant 0305] Writing is slower than speaking. When writing, I could not write out everything, so when speaking, I added more details. [Participant 0301]

In short, it seems that writing before speaking encourages more focus on form and lexical and structural variety than rehearsals.

Interview question 6: "If you are allowed some time to plan for another similar oral task in the future, how would you choose to plan your speaking?"

When asked about their favorite planning type, most interviewees of the three groups said that they would choose to write notes of key words or an outline with main ideas in phrases rather than write complete sentences. Eleven out of 13 participants (85%) who answered the question shared similar views that when looking at notes, speakers can elaborate. On the other hand, those who planned by writing claimed that writing in complete sentences was timeconsuming and they found it hard to remember all ideas because the story telling was too long. One interviewee suggested allowing more planning time (about 20 minutes) to write up the story. Another interesting explanation for not supporting writing out was that writing and speaking encourage attention to different aspects of language as reflected in the following example:

When writing out everything, we have to think in a different way because writing requires more attention to grammar, thus taking more time from exploring the details and searching for expressions than speaking. [Participant 0304]

However, one first-year student who confessed not being good at grammar and vocabulary liked writing out on paper and try to remember to retell the story. Another student said that if given planning time, he would search for sentences and ideas and try to remember the main ideas; if it is an important task, he would write out the whole thing.

### Results from qualitative analysis of the four special cases

An analysis of the oral and written planning data and the final oral narratives of four participants, one high achieving and one low achieving each from the P-W and P+W groups, revealed some noteworthy differences between planners.

For the P-W group, I chose two participants, one strong and one weak, coded as 0205 and 0210. Participant 0205 was one of the most fluent and accurate speakers of the group. His speech rate was 123.82 syllables per minute compared with the group average of 100.05 (SD = 16.64), and he made only 3.9 errors per 100 words while the group mean was 10.75 (SD = 4.12). During planning, he rehearsed two times with only several self-repairs and remarkably improved his overall complexity and subordination. An examination of his first oral narrative revealed that this participant created 15 simple T-units and only one complex T-unit with a total of 176 words (excluding dysfluencies), whereas his final performance had 14 simple T-units and 7 complex ones with 282 words in total. In other words, he spoke 11 words per T-unit at planning time 1 but increased the number of words up to 13.43 words per T-unit in his final narrative. Also, there was an increase in the number of clauses per T-unit in his narratives from the first time of planning (1.13 clauses/T-unit) to the final performance (1.38 clauses/T-unit). He seemed to pay more attention to expanding the sentences as can be seen through his two times of planning and final performance. In the two examples below, he described the two groups of boys. Every time he repeated the narrative, he added more modifying phrases (as in Example 1) and relative clauses (as in Example 2).<sup>2</sup>

Example 1

Planning time 1: *Three small boys standing in front of a bus.* 

 $<sup>^{2}</sup>$  The examples cited are taken from the raw oral data with all the dysfluencies underlined and pauses replaced with three dots (...).

Planning time 2:	One day three small boys in red shirts are waiting for a bus to
	come.

Performance:	One day three small boys in red shirts standing in front of a bus
	station to wait for the bus to come to get them to school.

Example 2

- Planning time 1: And <u>a group of another</u> a group of other ... four standing close to them.
- Planning time 2:There were other four boys in blue shirts who also come in line ...wait in line to get ... to the bus.
- Performance:And at the same time there were ... four boys who are older thanthe three ... younger boys, and they were in blue shirts.

Participant 0205 not only expanded the sentences but also made more connections between ideas by using subordinate clauses through the narrative planning and performance. This is illustrated by the following example.

- Planning time 1:And they are about to enter ... to get on the bus number twenty-six... All of the boys stand ... in line in order to get on the bus ... Butthe four boys in blue shirts enter first, ... and the remaining threeboys ... still standing outside the bus at three o'clock ... The threeboys in red shirts seem to be ... angry, ... and they are they keepwaiting for another bus to come.
- Planning time 2: The three small boys came first while the four boys came later ... They are waiting in line to get on the bus number twenty-six ... The four big boys play trick on the three younger, and they get on the

bus first while the three other boys ... couldn't. They seem to be surprised and ... angry.

Performance:All of the boys wait in line in order to come in order to get on the<br/>bus. The bus number twenty-six finally came. But the four boys<br/>who came later get on the bus first, while the three ... small boys<br/>still stay to wait for another bus to come ... The three boys seem to<br/>be very ... surprised and a little bit angry because they came first<br/>..., but they had to stay to wait for another bus to come.

The second participant of the P-W group (Participant 0210) selected for this analysis was one of the least accurate speakers of the group. Despite rehearsing the narrative three times and making many self-corrections, she still made nearly 20 errors per 100 words in her final narrative while the group mean was 10.75. In contrast to Participant 0205, who seemed not to pay attention to any grammatical point, Participant 0210 dispersed her attention over many linguistic forms such as verb tense, pronunciation, word choice, articles, quantifiers, and voice. This is noted by examining the corrections in her planning and performance in the following example.

Planning time 1:One day ... three small boy in red shirt are waiting were were<br/>waiting a bus ... waiting the bus ... Suddenly a truck ... run ...<br/>fastly and <u>mud</u> ... makes the boy make the boys ... muddy make the<br/>boy <u>muddy</u> ... muddy. After that ... the bus arrive ... and ... elder<br/>peop ... the bus with number thirty-six arrive and make the peop<br/>and <u>el</u> elder body hustle each other and ... make <u>one of the boys</u><br/>sfall fall fell one of the boy felt. At the end these small boy cannot<br/>go get on ... the bus.

- Planning time 2:One day three small boy in red T-shirts are waiting at the bus stop... On the road ... there <u>are</u> was a pile ... <u>a mud</u> of mud ... Therewas <u>many mud</u> ... there was there was much mud ..., and the truck... and a truck ... make ... three small boys ... dirty ... After that ...the bus ... the bus come, and everybody ... hustle to get on the buyt... with number thirty twenty-six, and they ... dominate three smallboy. And after that ... three small boy cannot ... get on the buytbecause of it is out of seak.
- Planning time 3: One day three small boy in red shirt are waiting at the bus stop ... with a lot of people. Suddenly <u>a truck ... drive fastly</u> ... a truck were driven fastly and <u>make ... the boy</u> make the boys muddy ...
  And ... after that ... <u>a buyts bus</u> a bus come. Everyone hustle each other to get on the bus, and these small boy cannot get on because the buyts was out of the seat.
- Performance:One day three small boy in red t-shirt ... as well as an many people<br/>are waiting were waiting at bus stop ... There were a ... were<br/>much mud on the road ... And a truck ... drive fastly so that it<br/>make it made three small boy muddy ... After that a buyt with the<br/>number thirty-six come. Everyone hustle each other and get on<br/>hustle each other to get on the bus. And three small boy cannot<br/>got on because of it's out it is out of the seat.

As partly seen in the above example, Participant 0210 made fewer corrections and pauses in her final performance than in her first rehearsal; however, her performance showed limited success in achieving better accuracy except some correct forms of past tense verbs (e.g., *are waiting*  $\rightarrow$  *were waiting*) and one quantifier (e.g., *many mud*  $\rightarrow$  *much mud*). She failed to pronounce the noun *bus* properly and dropped the plural ending *-s* (as in *boys*) in most cases though she attended to it at planning time 1 and 3.

Another difference between Participant 0205 and Participant 0210 lies in the complexity level of their narratives. While Participant 0205 gave a focus on the complexification of his language, Participant 0210 used only simple sentences at the first time of planning and then only one complex sentence in the two subsequent rehearsals. She retained this complex structure in her performance, but it was not error-free (e.g., *And three small boy cannot got on because of it is out of the seat*). In fact, she neither added many new ideas nor expanded the structures during planning; her narrative at the first planning time had 12 simple T-units with a total of 128 words (i.e., 10.67 words/T-unit), whereas her performance contained 9 simple T-units and 2 complex ones with 132 words in total (i.e., 12 words/T-unit). Thus, there was not much change in terms of overall complexity and subordination between planning and performance in the case of Participant 0210.

A closer comparison of the number of past tense verbs used in the narratives showed a remarkable dissimilarity between the two selected participants of the P-W group. Participant 0210 used fewer verbs in the past tense during performance (29%) than at planning time 1 (33%) while Participant 0205 increased the percentage of verbs in the past tense from 22% in planning to 59% in performance. It may mean that the stronger participant had more attention to and better control of the past tense usage than the weaker one.

For the P+W group, I also chose two participants, one high achieving and one low achieving, coded as 0308 and 0305. Participant 0308 was one of the most accurate speakers of

the group. His speech rate was 115.04 syllables per minute compared to the group average of 108.11 (*SD* = 19.83), but he made only 5.36 errors per 100 words while the group mean was 10.88 (*SD* = 4.27). One noteworthy similarity between his written and oral narratives is the high percentage of past tense verbs used (e.g., 93% for planning and 85% for performance). A more careful look at his writing showed that he used only one verb in the present tense and seemed to be aware of using the past tense to narrate the story by changing two verbs from the present to the past tense during planning. Apart from his attention to verb tense, he also made two corrections with word choice and one change with an expression in his writing. The fact that he achieved the highest score of the grammar pretest (77 correct answers out of 80 items) could also predict his great accuracy in both writing and speaking.

Regarding complexity, Participant 0308 carried the same structural variety in his writing over to his speaking; both of his written and oral narratives contained four different verb forms including simple past, past progressive, past passive, and past modal verbs. Although there was almost no gain in subordination since both his writing and speaking had similar clause/T-unit ratios (e.g., 1.36 for writing and 1.38 for speaking), an increase in the overall and phrasal complexity from his planning to speaking could be observed. For instance, his written narrative had 11 T-units and 15 clauses with a total of 119 words, whereas his final performance had 13 T-units and 18 clauses with 168 words in total. In other words, he wrote 10.82 words per T-unit but spoke 12.92 words per T-unit. Similarly, there were 7.93 words per clause in his writing, but his speaking had 9.33 words per clause. The participant's ability to expand the sentences can be illustrated as follows.

Planning:

When a bus numbered twenty-six came, the three boys were pushed back the queue by a man.

Performance: When a bus numbered twenty-six came, the three boys were waiting in a line but couldn't get on the bus because an aggressive man pushed them back.

In contrast to Participant 0308, Participant 0305 was one of the least accurate speakers of the P+W group. His speech rate was 95.43 syllables per minute, below the group average of 108.11, and he made 15.94 errors per 100 words, higher than the group mean of 10.88. During planning, Participant 0305 made four corrections with verb tense changing the verbs from the present to the past tense, two corrections with articles, and the other corrections with one preposition, one non-finite verb form, one verb in the passive voice, and spelling. It seemed that this participant gave more attention to verb tense than other grammatical features; however, he could use only 26% of verbs in the past tense in his oral narrative while 80% of the verbs in his written narrative were planned in the past tense.

In terms of complexity, there was also some evidence indicating that Participant 0305 could not use much of what he had planned in writing. For example, his oral narrative had less overall and phrasal complexity than his written narrative. He spoke 12.55 words per T-unit and 7.67 words per clause, whereas he wrote 13.20 words per T-unit and 9.43 words per clause. Moreover, there was no noticeable improvement in the text length of his writing and speaking since both modalities had a similar number of words, 132 words in the written narrative and 138 words in the oral narrative. Moreover, the structural variety decreased from his writing to his speaking. In his written narrative, there were three different verb forms including past simple, past progressive, and past modal verbs while there was only one simple past verb form in his oral narrative. However, there was an increase in the clause/T-unit ratio from his writing (1.40) to his speaking (1.64).

In general, it was evident that the two strong participants of both groups made fewer errors per 100 words and spoke more words per T-unit in their performance than the weak participants. It was also observable that the total number of words was greatly increased in the final narratives of the two high achievers, especially Participant 0205 of the P-W group, who could increase his text length from planning to performance by 106 words. Another noticeable finding was that the strong participants could use the past tense verbs they had planned better than the weak participants. The strong participants' ability to create longer oral texts and use more past tense verbs in them has also raised the question of whether the planning groups differ in expanding the text length and using the past tense verbs. Further one-way ANOVA analyses revealed that there was a significant medium effect for planning on the total number of words spoken in the final narratives, F(2, 87) = 12.80, p < .01,  $\omega = .46$ , and on the percentage of past tense verbs, F(2, 87) = 3.42, p < .05,  $\omega = .23$ . Tukey HSD comparisons located more specific differences between the groups: the P-W produced significantly more words per narrative (M =225.30, SD = 66.39) than the NP (M = 168.23, SD = 49.46) and the P+W (M = 161.80, SD = 161.8041.69) (p < .01), and the P+W used significantly more past tense verbs (M = 46.47, SD = 27.75) than the NP (M = 25.61, SD = 29.83) (p < .05).

### **CHAPTER 5: DISCUSSION AND CONCLUSION**

This chapter summarizes the major research findings following the order of the research questions and discusses them in light of theoretical frameworks and previous studies on pre-task planning. It also addresses some limitations of the study and makes suggestions for future research. The chapter concludes with the contributions of the present study to the field of teaching ESL and to SLA theory.

RQ 1: What effects does planning with and without writing have on the *fluency* of an L2 oral narrative?

In general, results show that both planning with and without writing has a positive impact on the fluency of L2 oral narratives. Significant differences between these two planning groups and the no-planning group were found in two aspects of fluency—speed fluency and breakdown fluency, but not in measures of repair fluency. More specifically, those who planned their oral narratives by writing out the story in complete sentences spoke significantly faster than those who were not allowed time to plan. Also, those who rehearsed their performance made significantly fewer pauses per minute of speech, especially fewer filled pauses, than nonplanners, while those who wrote before speaking paused in the middle of clauses significantly less than non-planners. However, there was no significant difference between two types of planning, with and without writing, in all dimensions of fluency.

Given a comprehensive set of measures for fluency, this study confirms the beneficial effect of planning on the oral fluency of narrators since the finding of the study is consistent with those of previous research (e.g., Crookes, 1989; Foster & Skehan, 1996; Gilabert, 2007;

Mehnert, 1998; Skehan & Foster, 2005; Tavokoli & Skehan, 2005; Wigglesworth, 1997; Yuan & Ellis, 2003). The finding lends support to both the Cognition Hypothesis (Robinson, 2001, 2005, 2007, 2011) and the Limited Capacity Hypothesis (Skehan, 1998; Skehan & Foster, 1999, 2001) regarding the role of planning time in helping the task performers achieve greater fluency.

A surprising finding of this study is that planning with writing assisted the participants in significantly increasing their speaking speed during the task performance compared with no planning while planning with rehearsal did not. Although the only study by Kawauchi (2005) that investigated writing as a planning activity did not find any significant difference between the effects of planning with writing and planning with rehearsal on the fluency, complexity, and accuracy of oral narratives, the evidence that writing can improve oral fluency has been documented in previous studies on oral fluency (Blake, 2009; Hardison, 2011; Payne, 2002). The interview data of the current study also provided more insights into how writing can scaffold speaking. Many interviewees of the P+W group admitted remembering what they had planned, resulting in faster speech rate and fewer unfilled pauses.

Another interesting finding of this study regarding pausing was that the P+W group paused at mid-clause significantly less than non-planners. This pattern of pausing is closer to that of native speakers because prior studies found that native speakers, who also pause, have pauses at the end of clauses rather than at mid-clause (Skehan, 2009; Skehan & Foster, 2005; Tavakoli, 2011). In fact, a quick look at the baseline data showed that 10 native speakers paused at midclause per minute of narrative (M = 1.22, SD = 1.35) less than the P+W (M = 3.30, SD = 2.05), the P-W (M = 4.51, SD = 2.77), and the NP (M = 5.11, SD = 2.05). This feature of fluency measurement is argued to be what distinguishes native and non-native speech (Skehan, 2009). Data from the interview also revealed that most participants had the biggest problem with

vocabulary search, so it is not surprising to observe that non-planners paused anywhere in their speech instead of at the end of clauses when they were not allowed time to access their lexicalized knowledge.

In short, planning with writing and rehearsal benefited learners' oral fluency differently. Writing before speaking is likely to help learners speak faster and pause less at mid-clause, while rehearsal may help reduce the number of fillers in the speakers' narratives.

RQ 2: What effects does planning with and without writing have on the *complexity* of an L2 oral narrative?

The current study gives evidence for a small-to-medium effect for pre-task planning on the complexity of L2 oral narratives. Both types of planning had a significant effect on phrasal complexity measured by the number of words per clause. However, planning without writing had significant effects on overall complexity measured by the number of words per T-unit and lexical variety measured by the percentage of content words, whereas planning with writing significantly affected lexical variety measured by the type-token ratio. Overall, there is evidence for the significant effect of pre-task planning on the complexity of oral narratives though there is again no significant difference between the two types of planning—with and without writing.

The finding of the effect of planning on the structural complexity and lexical richness of oral narratives in the present study may allow us to agree with the Cognition Hypothesis (Robinson, 2001, 2005, 2007, 2011), which claims that learners can pay more attention to specific linguistic aspects of a conceptually demanding task with the support of planning time, thus inducing greater complexity in L2 oral production. Ellis (2009) also hypothesizes that planning alone may not have any significant effect on the complexity of task performance, but it

should be considered in combination with task design. Therefore, the oral task for this study was designed to be cognitively demanding enough for the learners to direct their attentional and memory resources to lexical, morphological, and syntactic aspects of the L2 system required to convey the message. In fact, the instructions for the task emphasize that this is an interesting story and the speakers should retell it as detailed as possible. During the interview, many learners also mentioned that they paid attention to how to make the story interesting by varying structures and word choice or using more adjectives to express the feeling of the characters. They recognized that planning time was also a great opportunity for them to explore more details and connect them in a coherent way. The learners' attention to the guidance and care about the feeling of the possible listener may contribute to the increased complexity of their oral narratives.

An interesting finding in answering the second research question is that all the planners produced significantly more words per clause than non-planners while there was no significant difference in the number of words per T-unit in the oral narratives by the planners who wrote before speaking and the non-planners. As Norris and Ortega (2009) explained, mean length of clause fundamentally differs from mean length of T-unit in that the former is not affected by the addition of subordinate clauses, but influenced by that of adjectives, adverbs, prepositional phrases, or non-finite verb phrases, thus it serves as a specific measure that taps complexification at the phrasal level. For this reason, clause length seems to be a more sensitive measure for complexity.

Despite inconsistent results, many previous studies report a positive effect of planning time on the complexity of learners' productions (e.g., Ortega, 1999; Sangarun, 2005; Tavokoli & Skehan, 2005; Wigglesworth, 1997; Yuan & Ellis, 2003), and the finding of this study generally

confirms this effect of planning. Noticeably, those who planned by writing did not prove to outperform those who planned with rehearsal in the aspect of linguistic complexity in the oral narratives as I predicted following Blake (2009), Payne (2002), Weissberg (2000), and Williams (2008), who hypothesized that writing encourages attention to linguistic form and there may be a transfer of skills from writing to speaking.

RQ3: What effects does planning with and without writing have on the *accuracy* of an L2 oral narrative?

As predicted by the Cognition Hypothesis (Robinson, 2001, 2005, 2007, 2011), planning with and without writing will support greater accuracy in the L2 learners' oral narratives than no planning. Like the effect of planning on the complexity of oral production, the current study also found a significant effect of planning on the accuracy of the learners' oral narratives measured by the number of correct verb forms. This finding does not coincide with Yuan and Ellis' (2003) result though the two studies used the same picture set and the same measure (i.e., the percentage of correct verb forms). Yuan and Ellis found no statistically significant differences in both error free clause and percentage of correct verb forms. To explain this result, they drew on Skehan and Foster's (1997) study, claiming that "learners may be predisposed to use their planning time to pay attention to how to organize and encode the propositional content rather than for searching their linguistic repertoire to maximize accuracy" (p. 20). Another possible explanation they failed to discuss is that the range of language proficiency of the participants they recruited was rather large (i.e., a TOEFL score range of 373-520), while learner proficiency is an important factor that affects the ability to perform tasks fluently and accurately, using complex language.

A question may arise concerning why increases in accuracy were noted in verb forms rather than in articles and the plural –s. This can be explained considering the L1 and L2 background of the participants. Traditionally, among the grammatical points Vietnamese EFL learners are taught, verb forms are given stronger emphasis than others. Most of the grammar lessons repeat how to use verb tenses, aspects, voice, modals, and subject-verb agreement because these linguistic features are the main focus of written English tests and exams. With a strong explicit knowledge of verb forms, the participants of the study could have been aware of them, leading to more noticing during planning and performance and then achieving greater accuracy in using these forms. In contrast, Vietnamese learners of English often misuse English articles and drop the final consonants in their speech since the Vietnamese language has a much simpler article system and does not have fricatives like the plural –s in final position.

Another noteworthy finding of the present study is that planning with writing significantly reduced the number of lexical errors in the planners' oral narratives compared to no planning, while planning without writing did not. It suggests that writing may encourage more attention to linguistic form, leading to greater gains in accuracy (Kim, 2008; Williams, 2008). In fact, the interviewees from the P+W group revealed that writing before speaking helped them recognize the mistakes and they could correct them during performance.

## RQ4: What do learners attend to during planning?

A noticeable finding from planning data analysis is that the P-W planners spent more time focusing on meaning than focusing on form, whereas the P+W planners tended to focus more on form than on meaning, though the difference in the frequencies of focus on form and meaning is small. This finding satisfies the hypothesis that the P+W group will attend to

grammatical forms more than the P-W and NP groups, as predicted by previous research (Adams & Ross-Feldman, 2008; Weissberg, 2000; Williams, 2008). Since this is the first study that considers planning data as a great source to investigate what learners attend to during planning, any comparison with previous studies is impossible.

Planning data analysis also showed that focus on lexical forms accounted for 44% of total self-repairs in the P-W group and 37% in the P+W group. These percentages were higher than those of the other grammatical forms such as verb forms, articles, or prepositions. This finding is consistent with the prediction by Poole (2005) and Williams (1999) that all the planners will attend to more lexical forms than grammatical forms.

As predicted by Ortega (2005), the planners will attend to what they view as important (e.g., communication or accuracy) for task performance according to their individual preferences, language expertise, and interpretations of the task demands. In fact, interview data analysis indicated that the planners oriented their attention to many linguistic and non-linguistic aspects of their L2 production. However, they spent most of their time planning the content, searching for the right vocabulary, selecting the relevant grammatical structures, and monitoring their pronunciation. This seems to reflect the components of the L1 speech production process proposed by Levelt's (1989): conceptualization, formulation, articulation, and self-monitoring. According to Levelt, for L1 speakers, planning the message requires attention, but formulation and articulation are mainly automatic. Nonetheless, in L2 speech processing, many of the syntactic and phonological rules in L2 are not automatized (Kormos, 2011), and there are fluency vulnerability points (e.g., grammatical encoding, lexical retrieval, morpho-phonological encoding, phonetic encoding, and articulation) where L2 speakers' dysfluencies could result

from their difficulties in speech processing (Segalowitz, 2010). This explains why planning is beneficial for L2 speakers because it allows them time to circumvent these vulnerability points.

Finally, the analysis of the four special cases to gain more clues about the relationships between the planning data and the final narratives indicated that the types of planning are likely to benefit participants of different proficiency levels differently. The strong participants in both planning groups seemed to have better control of their linguistic accuracy and complexity, and thus could use much of what they had planned than the weaker participants. This may reflect the role of individual differences in the effects of planning, which needs further investigation as proposed by Ellis (2009). Nevertheless, the four participants' oral and written data showed no clear difference in the patterns of relationships between types of planning and the final oral performance. Based on how the four special participants planned their speech, further analysis of the text length was conducted, and it revealed that the P-W group produced significantly more words per pruned narrative than the other two groups, suggesting that text length or richness in ideas be a supplementary measure for the effect of rehearsal on L2 oral production. Moreover, the finding that the P+W group used significantly more past tense verbs in their speech than the NP group may encourage further research into the relationship between writing and speaking in consideration of other specific linguistic features.

### Limitations of the study and suggestions for future research

There remain several unavoidable limitations during the process of conducting this research that I can acknowledge.

First, the present study was mainly aimed to compare two types of planning: planning with writing and planning with rehearsal. Although both planning conditions showed a small-to-

medium effect on all the three aspects of L2 oral production, namely fluency, complexity, and accuracy, no significant difference between the effect of writing and rehearsal on oral narratives was detected. As observed in the data of four special participants, further research may look at the interaction between the type of planning and individual differences to see which type of individuals benefit most from the given planning condition. In addition, more measures such as text length, richness in details, or correct use of a specific linguistic aspect may be utilized.

Second, the study partly relied on the Cognition Hypothesis (Robinson, 2001) to explain the findings; however, the research design manipulated only one resource-dispersing variable (i.e., planning time) while assuming that the task was conceptually demanding without manipulating the resource-directing dimensions. As a consequence, future research may investigate the synergic effect of planning and another task complexity variable on the resourcedirecting dimensions.

Finally, the present research just looked at the immediate effect of rehearsal and writing on the fluency, complexity, and accuracy of L2 monologic narratives. Thus, there may be a need for conducting a longitudinal study to investigate whether any benefits gained from planning a task at one time carry over to a new task at a later time.

### Contributions of the study

This study provides more evidence for the beneficial effects of pre-task planning on the fluency, complexity, and accuracy of L2 oral performance. Results of the study are consistent with the predictions of both the Limited Capacity Hypothesis and the Cognition Hypothesis regarding fluency, but tended to lend more support to the Cognition Hypothesis in terms of complexity and accuracy. It is tenable that task complexity increased on the resource-directing

dimensions and decreased along the resource-dispersing dimensions can enhance all the aspects of language production. In addition, by employing both general and specific measures, the present study supports a comprehensive approach to understanding the effect of pre-task planning and proves that specific measures are more sensitive than general ones in measuring fluency, complexity, and accuracy.

This study does not only further our understanding of the effect of pre-task planning in planning research but also confirms its important role in task-based language teaching. Planning in any form, with writing or rehearsal, can benefit L2 learners at all levels depending on the task demands. The noninteractive nature of monologic narratives may assist L2 learners better because they can allocate full attention to using new forms or restructuring their interlanguage without being affected by interactive variables, thus improving their fluency, complexity, and accuracy at the same time.

The current study also adds to the literature on the oral-literate connection by providing more empirical evidence for the effect of writing on speaking while many researchers tend to look at the relationship between two modalities in the other direction, the effect of speaking on writing. The findings are of great significance to formal EFL instruction, where students are more exposed to the written form of the language than the spoken one, because they help supplement the classroom activities to transfer the students' skills from one modality to another.

APPENDICES

# Appendix A: Grammar pretest

For questions 1-60, you are to choose the <u>one</u> word or phrase, marked A, B, C, or D, that best completes each sentence. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter of the answer you have chosen.

1.	"Hurry up! We're waiting for you. What's taking you so long?" "I for an important phone call. Go ahead and leave without me."					
	A. have waited	B. will wait	C. am waiting	D. wait		
2.	A fortune-teller predicted		inherit a lol of money before the end of the			
	A. that I would	B. that I	C. what I will	D. what I		
3.	After nodding off.	to 45 minutes of an	extremely boring speech, I found myself			
	A. was listening	B. having listen	C. listening	D. I listen		
4.	Ann quit her job at the advertising agency, surprised everyone.					
	A. which	B. that	C. who	D. that it		
5.	Brian used to be an active person, but now he has to limit his activities problems with his health.					
	A. because	B. because of	C. although	D. in spite of		
6. A minor earthquake occurred at 2:07 a.m. on January 3. Most of the people in t at the time and didn't even know it had occurred until the new morning.						
	A. slept	B. had slept	C. were sleeping	D. sleep		
7.	Carrots raw provide the greatest nutritional value.					
	A. are eaten	B. eaten	C. which eat	D. eating		
8.	Dr. Sales is a person A. in whom I don't have much confidence B. in that I don't have much confidence C. whom I don't have much confidence in him D. I don't have much confidence					
9.	Had you told me that A, don't believe	this was going to happ	pen, I B. can't believe	it.		
	C. hadn't believed		D. would never have believed			
10.	I am tired of your rud	le behavior. I'm not go	bing to	_ it any more.		
	A. put away	B. put on	C. go over to	D. put up with		
11.	I don't understand ho	w John could have ma	de	in judgment.		
-----	-------------------------------------	---------------------------	-----------------------------------	---------------------		
	A. such big mistake		B. such a big mistake			
	C. so a big mistake		D. so big mistake			
	-		-			
12.	I hadn't expected Jam	nes to apologize, but I l	had hoped	·		
	A. that he would call	me	B. him calling me			
	C. him to call me		D. that he will call me	e		
13.	I have always wanted	to visit Paris,	of Fran	ce.		
	A. is the capital		B. which the capital is	S		
	C. that is the capital		D. the capital			
14.	I talked to Bob two w	eeks ago. I thought he	wanted to know about	my cat, but I		
	misunderstood him. H	He asked me where	, not r	ny cat.		
	A. is my hat	B. my hat was	C. my hat is	D. was my hat		
15.	I was enjoying my bo	ook, but I stopped	a progra	m on TV.		
	A. to read to watch		B. reading for watch			
	C. to read for watchin	ng	D. reading to watch			
16.	Dry farming is a type	of agriculture used in	areas	less than 20 inches		
	of rainfall.					
	A. there are	B. in which is	C. where there is	D. which has		
17.	I wish that you	such a lo	t of work, because we	know that you would		
	have enjoyed the part	у.				
	A. hadn't had	B. hadn't	C. didn't have had	D. hadn't have		
10						
18.	I can't	the answer to the	is math problem.	-		
	A. figure out	B. help out	C. take out	D. try out		
10	<b>-</b>					
19.	In many ways, riding	a bicycle is similar to	·································			
	A. the driving of a car $\tilde{a}$	r	B. when you drive a c	car		
	C. driving a car		D. when driving a car			
•	<b>-</b>					
20.	It is gravity	objects t	oward the earth.			
	A. pulling	B. that pulls	C. to pull	D. what pulls		
0.1		· 1 TT . 11				
21.	Mahfuz has an active	mind. He can retell m	emories	anybody else in		
	the classroom.	<b>D</b>		<b>D</b>		
	A. as vivid as	B. much vivid than	C. more vividly than	D. too vivid as		
~~						
22.	Mr. Smith dropped	this morning. I	He wanted to borrow so	ome milk me.		
	A. 1n - 0t	B. by - from	C. off - to	D. at - with		

23.	My family loves this	house. It	the family home ever since my	
	grandfather built it 60	) years ago.		
	A. has been	B. was	C. is	D. will be
24.	My children enjoy	to stay up late	when there's somethin	g special on TV.
	A. being allowed	B. to be allowed	C. allowing	D. to allow
25.	My writing has impro in writing.	oved a lot in this class.	. All the students	do well
	A. whom Mr. Davis t C. that Mr. Davis tead	eaches them ches them	B. which Mr. Davis t D. Mr. Davis teaches	teaches
26.	Not wanting to be late my bus.	e for my first day of c	lass,	to school after I missed
	A. therefore, I ran	B. because I ran	C. I ran	D. so I ran
27.	People who exercise A. who doesn't	frequently have greate B. that doesn't	er physical endurance the C. which don't	han those D. who don't
28.	Scientists	caves and the living t	hings in them are calle	d speleologists.
	A. studying	B. have studied	C. do a study of	D. study
29.	My mother always m	ade me	my hands before e	very meal.
	A. washing	B. to wash	C. washed	D. wash
30.	Sometimes very your many believe that dra	ng children have troub Igons actually exist.	le f	act from fiction, and
	A. separating	B. to separate	C. for separating	D. separate
31.	The bananas wouldn'	t have spoiled unless	the weather	hot.
	A. was being	B. was	C. has been	D. had been
32.	The Chicago bus is p	arked at	•	
	A. the lane two	B. the two lane	C. lane two	D. second lane
33.	The child's arm was	swollen because he	by a	bee.
	A. had being stung	B. had stung	C. had been stung	D. stung
34.	The first time I went	swimming in deep wa e learned to stay afloa	ter, I sank to the bottor t, I feel better about the	n like a rock. e water, but I still can't
	A. As soon as	B. The first time	C. When	D. Now that
35.		, Ruggiero Ricci wa	as considered a great vi	olinist.
	A. At six years old		B. At the age of six	
	C. When age six		D. When he was six	years aged

36.	. The man died because medical help was not summoned. A doctor should immediately.			
	A. have called	B. have been called	C. been called	D. called
37.	In the past, families it to school there, got r eventually died there	in North America were narrried there, brought	much closer. People g a	grew up in a town, went family there, and
	A. together with	B. up	C. with	D. in
38.	The power lines outs	ide my house were dar nem to a safer place.	ngerous. I finally got th	e power company
	A. move	B. to move	C. moving	D. moved
39.	There were	participants	at the conference than	we had predicted, so
	A. so many	B. too many	C. many more	D. much more
40.	"I was" "Her boss made her	that Jan couldn't c work overtime."	come to the party."	
	A. disappointing	B. disappointedly	C. disappointed	D. disappoint
41.	This is the woman	the artist	said posed as a model	for the paiting.
	A. which	B. who	C. whom	D. whose
42.	This machine is very	dirty. It needs	·	
	A. being cleaned	B. to clean it	C. be cleaned	D. cleaning
43.	Timmy doesn't do w	ell in school	his inability	to concentrate on any
	one thing for longer	than a minute or two.	C horrison	D 41 f
	A. due to	B. as	C. because	D. therefore
44.	44. "Mr. Wright, can you give me a little extra help typing some letters today?" "Sorry, can't. The boss has an urgent report for me to write. She demanded that it on her desk by 5 P M today."			rs today?" "Sorry, I d that it
	A. was	B. would be	C. is	D. be
45.	and listen.	saying was so importa	nt that I asked everyon	e to stop talking
	A. The woman was		B. What the woman	was
	C. That the woman w	vas	D. What was the wor	nan
46.	"The students all we "I heard it was really	nt to the circus yesterd	ay."	
	A. amused	B. amuse	C. amusing	D. amusingly

47.	She	her back while she	was playing volleyball	l last week in the game
	against South City Co	ollege.		
	A. hurted	B. hurt	C. was hurting	D. had hurted
48	"You look nervous"	"This thunder make	s me	"
40.	A scary	B scare	C scaring	 D_scared
	rt. seary	D. searc	C. seamg	D. seared
49.	To answer accurately	is more important that	n	<u> </u>
	A. a quick finish		B. to finish quickly	
	C. finishing quickly		D. you finish quickly	
50	Today many serious	childhood diseases	by early	immunization
50.	A. are preventing	B. can prevent	C. can be prevented	D. prevent
	I I I B	I I I I I I I I I I I I I I I I I I I	I	Ī
51.	Tom is sitting at his d He	esk. He's reading his c	chemistry text because	he has a test tomorrow.
	A. must be studying	B. should be studying	g C. will study	D. could study
50	When I kept getting u	numerical colles. I colled	the phone company of	ad had my phone
52.	number	ilwaineu cans, i caneu	the phone company a	iu nau my phone
	A change	 B_changing	C to change	D changed
	in enunge	Dremanging	et to enunge	Dienangea
53.	You should learn how emergency situation it	y to change a tire on yo f necessary.	our car	_ you can handle an
	A. so that	B. when	C. for that	D. therefore
54.	4. The plane's departure was delayed because of mechanical difficulties. When the weary passengers finally boarded the aircraft, many were annoyed and irritable because they in the airport for three and a half hours.			ies. When the weary itable because they
	A. are waiting		B. had been waiting	
	C. have been waiting		D. were waiting	
55.	"Were you pleased wi	ith the translation?" _	"Yes, the job was	.,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
	A. amazing satisfactory		B. amazingly satisfactory	
	C. amazingly satisfact	torily	D. satisfactory amazin	ng
56	The Smiths were a ve	ry happy family. The a	hildren got	each other
50.	A. together with	B. along with	C. through with	D. up
	8	6		
57.	There is so	heat in Long	don this summer.	
	A. few	B. any	C. many	D. much
58	"Why are you so late?	)"		
50.	"I my arc you so fale?	v aunt to the airport T	he traffic was terrible '	,
	A. could take	B. must have taken	C. should take	D. had to take

59. \_\_\_\_\_\_ powerful member of the entire cat family is the tiger.
A. The more B. More C. The most D. Most
60. "Hello! Alice? This is Jeff. How are you ?"
 "Jeff? What a coincidence! I \_\_\_\_\_\_ about you when the phone rang."
 A. was just thinking B. just thought

A. was just thinkingEC. have just been thinkingE

D. was just thought

For questions 61-80, you are to read the following passage and choose the <u>one</u> article (a, an, the, or zero article X) that best completes each blank. Then, on your answer sheet, find the number of the blank and fill in the space that corresponds to the letter of the answer you have chosen.

I could tell by (61) \_\_\_\_\_\_ townhall clock that I was late, so I decided to catch (62) \_\_\_\_\_\_ bus. It was a beautiful day; the sun was shining and there was (63) \_\_\_\_\_\_ very little wind. I turned the corner, and walked down the main street.

A couple of minutes later, I heard (64) \_\_\_\_\_ noise, and (65) \_\_\_\_\_ man wearing (66) \_\_\_\_\_ grey leather jacket ran past me. At first, I thought he was trying to catch (67) \_\_\_\_\_ bus which was waiting at the bus stop, but then (68) \_\_\_\_\_ policeman appeared, running at some speed. He was obviously chasing (69) \_\_\_\_\_ man in (70) \_\_\_\_\_ leather jacket; and he was joined by another policeman, who was talking rapidly into a hand-held radio.

All three disappeared into (71) \_\_\_\_\_\_ crowd of people. My bus arrived, and I got on. As (72) \_\_\_\_\_ bus drove down the road, I saw (73) \_\_\_\_\_ man again, walking casually through (74) \_\_\_\_\_ crowd with his coat over his shoulder. I could also see (75) \_\_\_\_\_ second policeman, still talking into his radio. He was describing (76) \_\_\_\_\_ man who no longer existed, (77) \_\_\_\_\_ man wearing (78) \_\_\_\_\_ jacket and running furiously while (79) \_\_\_\_\_ real criminal (if he was (80) \_\_\_\_\_ criminal) walked slowly and casually into the station.

61.	A. a	B. the	C. X
62.	A. a	B. the	C. X
63.	A. a	B. the	C. X
64.	A. a	B. the	С. Х
65.	A. a	B. the	С. Х
66.	A. a	B. the	С. Х
67.	A. a	B. the	С. Х
68.	A. a	B. the	С. Х
69.	A. a	B. the	С. Х
70.	A. a	B. the	С. Х
71.	A. a	B. the	С. Х
72.	A. a	B. the	С. Х
73.	A. a	B. the	С. Х

74.	A. a	B. the	С. Х
75.	A. a	B. the	C. X
76.	A. a	B. the	С. Х
77.	A. a	B. the	С. Х
78.	A. a	B. the	C. X
79.	A. a	B. the	C. X
80.	A. a	B. the	C. X

#### Appendix B: Pronunciation pretest

Please read aloud the following paragraph at your normal speed and as carefully as you can.

My Journey to the U.S.

My name is Viet, and I'm from Vietnam. As a high school student, I always *dreamed* of going to the *United States*. In English *classes* at school, my teacher *talked* about the country a lot. I *was fascinated* with the beautiful *cities* and friendly people that my teacher told us. I *searched* for more information about the country and *read* many English *stories* about it, but I couldn't understand them all because my English *was* very *limited*. Whenever I got stuck in understanding the *structures* or vocabulary in the *stories*, I often *asked* my teacher for help. She *was* the first teacher who *inspired* me with American culture and higher education. When I *received* a scholarship to study abroad, I knew I *reached* a turning-point in my life and then left everything behind to pursue my American dream.

*Note*. The version for participants does not have the past tense verbs or adjectives with *-ed* endings and plural nouns with *-s* or *-es* endings put in bold face type and in italics.





#### Appendix D: Task instructions in English and in Vietnamese

#### English Version

#### Condition 1: No Planning (NP)

Here is a set of 6 pictures already ordered from 1 to 6. These pictures tell you an interesting story. Now I would like you to retell this story in English. Please tell the story as detailed as possible as if you were telling it to a friend who has never heard the story before so that he/she can visualize what happened. You have 5 minutes to tell the story, and your story should start with: *One day, three small boys in red shirts* .....

#### *Condition 2:* Planning without Writing (P-W)

Here is a set of 6 pictures already ordered from 1 to 6. These pictures tell you an interesting story. Now I would like you to retell this story in English. Please tell the story as detailed as possible as if you were telling it to a friend who has never heard the story before so that he/she can visualize what happened. You have 5 minutes to tell the story, and your story should start with: *One day, three small boys in red shirts* .....

However, before telling the story, you will have 10 minutes to prepare what you are going to say. You should speak out loud in full sentences whatever you are preparing. Please do not take any notes. In 10 minutes, you will begin to retell the story.

*Condition 3:* Planning with Writing (P+W)

Here is a set of 6 pictures already ordered from 1 to 6. These pictures tell you an interesting story. Now I would like you to retell this story in English. Please tell the story as detailed as possible as if you were telling it to a friend who has never heard the story before so that he/she can visualize what happened. You have 5 minutes to tell the story, and your story should start with: *One day, three small boys in red shirts* .....

However, before telling the story, you will have 10 minutes to prepare what you are going to say. You should write out the entire story in full sentences on the given sheet of paper. In 10 minutes, your writing will be taken away, and you will begin to retell the story.

### Vietnamese Version

### Điều kiện 1: Không chuẩn bị

Đây là một bộ 6 bức tranh đã được sắp xếp theo thứ tự từ 1 đến 6. Những bức tranh này kể lại một câu chuyện thú vị. Bạn hãy kể lại câu chuyện này bằng tiếng Anh. Hãy kể câu chuyện thật chi tiết như thể bạn đang kể cho một người bạn chưa từng nghe đến câu chuyện sao cho người này hình dung được chuyện gì đã xảy ra. Bạn có 5 phút để kể, và bạn nên bắt đầu câu chuyện bằng: *One day, three small boys in red shirts* .....

### Điều kiện 2: Chuẩn bị không có viết

Đây là một bộ 6 bức tranh đã được sắp xếp theo thứ tự từ 1 đến 6. Những bức tranh này kể lại một câu chuyện thú vị. Bạn hãy kể lại câu chuyện này bằng tiếng Anh. Hãy kể câu chuyện thật chi tiết như thể bạn đang kể cho một người bạn chưa từng nghe đến câu chuyện sao cho

người này hình dung được chuyện gì đã xảy ra. Bạn có 5 phút để kể, và bạn nên bắt đầu câu chuyện bằng: *One day, three small boys in red shirts* .....

Tuy nhiên, trước khi kể chuyện, bạn có 10 phút để chuẩn bị những gì định nói. Bạn có thể nói to lên những gì bạn đang chuẩn bị. Xin đừng ghi chú. Sau 10 phút nữa, bạn sẽ bắt đầu kể lại câu chuyện.

## Điều kiện 3: Chuẩn bị kèm theo viết

Đây là một bộ 6 bức tranh đã được sắp xếp theo thứ tự từ 1 đến 6. Những bức tranh này kể lại một câu chuyện thú vị. Bạn hãy kể lại câu chuyện này bằng tiếng Anh. Hãy kể câu chuyện thật chi tiết như thể bạn đang kể cho một người bạn chưa từng nghe đến câu chuyện sao cho người này hình dung được chuyện gì đã xảy ra. Bạn có 5 phút để kể, và bạn nên bắt đầu câu chuyện bằng: *One day, three small boys in red shirts* .....

Tuy nhiên, trước khi kể chuyện, bạn có 10 phút để chuẩn bị những gì định nói bằng cách viết toàn bộ câu chuyện ra giấy bằng câu hoàn chỉnh. Sau 10 phút nữa, tôi sẽ thâu bài viết, và bạn sẽ bắt đầu kể lại câu chuyện.

## Appendix E: Interview questions

- 1. What were you thinking about while planning? What did you plan?
- What were you most concerned about during planning? Please explain the reasons for your concerns.
- 3. Do you think you performed the oral task successfully? Why or why not?
- 4. Could you use what you had planned in your narrative? Why or why not?
- 5. What are the benefits of the type of planning you experienced?
- 6. What are the difficulties of the type of planning you experienced?
- 7. If you are allowed some time to plan for another similar oral task in the future, how would you choose to plan your speaking?

# Appendix F: Coding guidelines for fluency measures

Categories	Explanations	Examples
Speed fluency		
Speech rate	Number of syllables per minute. Divide the total number of syllables in a speech sample by the total number of seconds used to produce the sample (including pause time), and then multiply the result by 60. For unfilled pauses over 3 seconds, include 3 seconds of pause in the calculation and exclude the remainder (Riggenbach, 1991). A one-to-three second pause is marked with <i>pause</i> in brackets.	Suddenly a truck coming a truck is coming and [pause] and the truck make they dirty with the water on the street. Total number of syllables = 27 Total time used = 14 sec $(27 \text{ syl} / 14 \text{ sec}) \times 60 = 115.71 \text{ syl/min}$
Breakdown flue	ency	
Pauses	Number of pauses per minute. Divide the total number of pauses in a speech sample by the total number of seconds used to produce the sample, and then multiply the result by 60.	One day <i>uh</i> three small boys in <i>uh</i> red shirt <i>uh uh</i> go [pause] go along a street. <i>Uh</i> suddenly a truck [pause] the truck <i>uh</i> go by [pause].
	Count only pauses of 1 second or more. Calculate unfilled and filled pauses separately.	3 unfilled pauses (1 end-clause and 2 mid-clause) + 6 filled pauses
	Unfilled pauses include pauses at the end and in the	Total time used = $13 \text{ sec}$
	middle of clauses.	(9 pauses / 13 sec) $\times$ 60 = 41.54
	Filled pauses include <i>ah</i> , <i>um</i> , <i>uh</i> , <i>well</i> , <i>you know</i> , etc.	pauses/min
Repair fluency		
Dysfluencies	Number of dysfluencies per minute. Divide the total number of dysfluencies in a speech sample by the total number of seconds used to produce the sample, and then multiply the result by 60.	(total number of dysfluencies / total time in sec) $\times$ 60 = dysfluencies/min
	Repetitions are syllables, words, phrases, or clauses that are immediately repeated with no modification (verbatim repetition).	<i>There are</i> there are many people. And when the <i>pa pa</i> passengers get on the bus <i>one</i> one boy seems to be fall into onto <i>the</i> the street. [5 repetitions]
	Exclude repetitions for rhetorical effect.	They got <i>really</i> really wet.
	Reformulations/self-corrections are words, phrases, or clauses that are repeated with some modification to morphology, syntax, pronunciation, or word order. When a reformulation contains a pause or a repetition, count the pause or the repetition separately.	They <i>decide</i> decided to <i>catch a bust</i> catch a bus. A bus running and they cause and <i>the</i> the bus causes the boys <i>wet to</i> to get wet. [4 self-corrections; 2 repetitions]
	Replacements are lexical items that are immediately substituted for another.	Uh suddenly a truck uh the truck go by and hit uh and hit pool <i>some pool</i> uh some mud on their clothes. [1 replacement]
	False starts are utterances that are abandoned before completion and that may or may not be followed by a reformulation.	Um <i>they go to</i> um finally the bus come, but the but they aren't accepted. [1 false start]

# Table 17: Coding guidelines for fluency measures

# Appendix G: Coding guidelines for complexity measures

Categories	Explanations	Examples
Structural com	plexity	
Overall complexity	T-unit length or mean length of T-unit. Divide the number of words by the number of T-units in each pruned narrative.	One day three small boys in red shirt go along a street. Suddenly a truck go by. 17 words / 2 T-units = 8.5 words/T-unit
	A T-unit consists of an independent clause	The small boys got on the bus. [1 T-unit]
	with or without dependent/subordinate clauses (i.e., noun, adjective, and adverbial clauses).	They noticed that the bus that they were allowed to get on had accidentally broken down. [1 T-unit; 1 independent clause and 2 dependent clauses]
	When there is a grammatical subject deletion in a coordinate clause, count the entire sentence as 1 T-unit.	A truck drove by and splashed water all over the boys. [1 T-unit]
	For direct speech, count the reporting clause and all the T-units (in quotation marks) as individual T-units.	He said, "I'm sorry, but you cannot get on." [3 T- units; 3 clauses]
Phrasal complexity	Clause length or mean length of clause. Divide the number of words by the number of clauses in each pruned parrative	They noticed that the bus that they were allowed to get on had accidentally broken down.
	A loss in each planed narrative.	16  words / 3  clauses = 5.33  words/clause
	A clause consists of an overt subject and a finite verb phrase.	The driver solved the house to wait [1 cloues]
	Count on importive of 1 clouds	Stav away from ma [1 clause]
Subordination	Clause/T-unit ratio. Divide the number of clauses by the number of T-units in each pruned narrative.	One day three small boys were waiting for the bus when a truck came by and splashed them with water from the road. When more people arrived for the bus and the bus arrived, the boys were pushed to the back of the bus line because they were dirty. 6 clauses / 2 T-units = 3 clauses/T-unit
Structural variety	Number of different verb forms in each pruned narrative, including past tense aspects (e.g., simple past, past progressive, and past perfective), modality (e.g., present modals like <i>should</i> , <i>have to</i> , and past modals like <i>should have</i> , <i>had to</i> ), and voice (e.g., passive voice in the past)	One day three small boys <i>were standing</i> at the bus stop to wait for a bus. Then the bus arrive[d]. They <i>wanted</i> to get on, but they <i>were push</i> behind by the bigger boys in blue shirts. They <i>couldn't get on</i> because the bus <i>was</i> full. They <i>had to waited</i> to three-thirty for the next bus [4 different verb forms: simple past, past progressive, past passive voice, and past modal]
Lexical richnes	S	
Lexical variety	Type-token ratiothe ratio in percent between the different lexemes and the total number of lexemes within the first 100 words of each pruned narrative	As the bus twentysix arrive at the busstop the people queueup behind the boy kickout them and geton the bus before the three small boy (18 different lexemes / 25 lexemes) ×100% = 72%
Lexical density	Percentage of lexical words in each pruned narrative (i.e., nouns, verbs, adjectives, and adverbs)	The three small boys in red shirts wave their hands to say goodbye to the people in the bus twenty-six. (12 lexical words / 20 lexemes) $\times$ 100% = 60%

# Table 18: Coding guidelines for complexity measures

# Appendix H: Coding guidelines for accuracy measures

Table 19: Coding guidelines for accuracy measures
---

Categories	Explanations	Examples		
Errors per 100	Errors per 100 words			
	Divide the total number of errors by the total number of words in each pruned narrative and then multiply the result by 100%.	One time there <u>was</u> three <u>boy</u> in red <u>was</u> waiting at the bus stop. Suddenly <u>the</u> big <u>trunk</u> <u>come</u> and <u>make (<i>missing 'the'</i>)</u> three <u>boy</u> dirty. (9 errors /24 words) $\times$ 100% = 37.5%		
	Count consistently repeated errors due to wrong assumptions only once.	the bus twenty-six (one example of the repeated errors)		
Correct verb fo	rms			
	The percentage of correct verb forms used in obligatory situations. Identify all the verb forms in relation to tense, aspect, voice, modality, and subject-verb agreement. Then, divide the total number of correct verb forms by the total number of verbs used in each pruned narrative, and then multiply the result by 100%.	One day three small boys <i>were standing</i> at the bus stop to wait for the bus. The bus <i>is coming</i> . At the bus stop there <i>were</i> many people. The people one by one <i>went</i> in the bus. But our boys <i>standing</i> behind many other people. (3 correct verb forms / 5 verbs) $\times$ 100% = 60%		
	Types of verb-form e	errors		
Tenses and aspects	Incorrect Misformed Incorrect or misformed	The driver said that the bus <u>is</u> full. The bus <u>didn't had</u> enough room. As they were waiting there, four other boys		
	Past tense verbs incorrectly inflected	<u>were come.</u> They <u>standed</u> there for half an hour.		
	Unnecessary/unmotivated shift in tense. Tense errors should be identified based on the discourse context or the tense sequence within the same T-units or in independent clauses joined by <i>and</i> , <i>but</i> , <i>or</i> , <i>so</i> .	The bus of the mean boys broke down on the way. The little boys <u>laugh</u> happily when they got to the destination sooner. Maybe they didn't have enough money, so they <u>can't</u> get on the bus.		
	However, don't count the historical present tense verbs in the narrative as errors.	There are some people who are waiting at the bus stop. (no error)		
	Don't count the simple present for evaluation or comments as an error.	So the bus of the bigger boys broke down in the middle of the road. It's interesting to see this. (no error)		
	Don't count it as an error if you think the simple past is being used and the speaker's pronunciation of $-ed$ or $-d$ is not clearly heard due to assimilation.	The driver close(d) the door of the bus and drove away. That bus happen(ed) to be the bus of the boys in the blue shirts.		
Voice	Incorrect Misformed	Three children <u>don't allow</u> to take the bus. They don't be accepted to get on the bus.		
Modality	Wrong modals Wrong form of modals Wrong form of the main verb following the modal	The small boys were not allowed to get on, so they <u>could</u> wait for another bus. The driver said that another bus <u>will</u> come soon. They can got on the bus.		

Table 19 (cont'd)

Subject-verb	Lack of subject-verb agreement.	Unluckily, the bus were full at that time.
agreement	Count this case as an error if there is evidence that the historical present tense is being used.	The bus driver <u>say</u> it's too late.
Target-like use	of articles	
I ur gei-like use	Divide the number of accurately supplied	One day three small have in red shirts were
	articles by the number of obligatory contexts	waiting at <i>the</i> bus stop. <i>The</i> street was very wet
	and inappropriately supplied articles in each	because of <i>the</i> rain. Suddenly <u>the</u> big truck
	pruned narrative, and then multiply the result	passed by and made <i>the</i> boys dirty. When <u>the</u>
	by 100%.	came, ( <i>missing 'the'</i> ) three boys could not get
		on <i>the</i> bus.
		[5 accurate suppliances / (7 obligatory contexts
		+ 1 inappropriate suppliance)] $\times$ 100% = 62.5%
Target-like use	of the plural – <i>s</i>	
	Divide the number of accurately supplied	One day three <u>boy</u> in red <i>shirts</i> want to make a
	and inappropriately supplied plural –s in each	the bus, they are pushed by some older <i>peoples</i>
	pruned narrative, and then multiply the result	( <i>inappropriate suppliance</i> ). And they can't take
	by 100%.	the bus because it is full of <i>passengers</i> . After
		thirty <u>minute</u> they take another bus.
		+ 1 inappropriate suppliances / (4 obligatory contexts + 1 inappropriate suppliance)] $\times$ 100% = 40%
Lexical errors		
	The raw number of lexical errors in each pruned	I narrative. Include lexical errors related to
	idioms, and expressions that interfere with the c	ammatical form, word order, collocation,
	Types of lexical en	rors
Word	Count serious deviations in pronunciation or	struck instead of truck
pronunciation	completely unintelligible words.	<i>pus</i> instead of <i>bus</i>
		sheets instead of shirts
	Don't count slight mispronunciation of a word	alway
	as an error.	sudden <u>t</u> ly
Word	Mixing up words that sound similar	effect instead of affect
meaning	Using the wrong shade of meaning	watch instead of clock
	Using a word with a completely wrong	department instead of pavement
	meaning	
Word form	Wrong word forms (corrections provided in parentheses)	They look very <u>hurry</u> . (hurried)
XX7 1 1		The boys are very <u>surprising</u> . (surprised)
Word order	wrong word order	in the bus.
Collocation	Wrong collocation	to make a favor instead of to do a favor
	wrong conocation	to make a javor mistead of to do a javor

Table 19 (cont'd)

Awkward	Awkward phrasing - any form of non-native-	to make a travel to travel around the city
expressions	like phrasing that may sound unnatural or confusing to native speakers of English	A truck makes the dirty water fly into the people's bodies.
		One day three small boys in red shirts escape their family for playing something.
Other types of e	errors (Count each following case as 1 error.)	
Non-finite	Wrong non-finite verb forms including	The driver didn't let them to get on. (get)
verb forms	inappropriate infinitives, present participles,	They were not permitted getting on the bus (to
	and past participles (corrections provided in	get)
	parentheses). These non-finite verbs are not	There are many people go to work. (going)
	the main verbs in each clause.	The driver had the wheel fix. (fixed)
Ouantifier-	Lack of agreement between quantifiers (e.g.	We can see many passenger at the bus station
noun	much, many, this, these, and so on) and nouns	One children was crying
agreement		one <u>ennaren</u> was erynng.
Pronoun	Vague pronoun reference	The little boys were waiting for the bus with
reference		many other people. <u>They</u> were at the end of the
		line.
Case	Wrong subject or object case	The truck made <u>they</u> dirty.
Preposition	Wrong preposition	There are three boys waiting <u>on</u> the bus stop.
Missing	Count all missing words, including articles,	The boys were standing at ^ bus stop. (the)
words	prepositions, verbs, auxiliaries, subjects,	Many people are waiting ^ the bus. (for)
	provided in parentheses).	They think they will ^ late for school. (be)
Extraneous words	Count all extraneous words as errors.	Three boys stand on the pavement <u>but</u> with others.
		There are many persons are at the bus stop.
		Among of them were a group of students.
	Don't count the pronoun that follows a	Some boys they are waiting for the bus.
	topicalized noun phrase as an extraneous	
	word.	
Sentence	If the verb or copula in a sentence is missing	They ^ on the way to school.
Iragments	(^), count the sentence as 1 1-unit with an error	
	If a noun phrase or a subordinate clause is	The bus is a number 26. When all the people
	standing alone, attach it to the preceding or	get on the bus. Now they can get on the bus.
	following T-unit as appropriate and count it as	
	an error.	
	Count an incomplete clause as an error.	The three small boys are at the end of the bus
		line, <u>so they can't.</u>
Dangling	Words, phrases, or clauses that are not close to	Looking down the street, it's dirty and muddy.
modifiers	the word they describe or relate to.	

Notes:

- 1. The underlined is incorrect.
- 2. In cases of fine decisions of appropriacy, no error was recorded.
- 3. Do not count any unfinished sentence at the end of a narrative.
- 4. As the above types of errors can be interrelated, don't double penalize for a word/phrase that relates to two errors (e.g., In the sentence *"They were not permitted going on the bus"*, the word *"going"* should be

corrected as "to get". If so, it can be categorized into both wrong verb forms and wrong words/phrasal verbs).

5. Be lenient with those errors that violate prescriptive grammar rules but are acceptable (e.g., *That's him* instead of *That's he.*)

### Appendix I: Data samples

## Speaking Sample from the Planning-without-Writing Group

One day three small boys in red shirts <u>was</u> were <u>standing</u> waiting for a bus at a bus station. While they were waiting for a bus, a trunk pass by and made them dirty. <u>Far</u> far away there were four boys in blue shirt coming [pause]. When a bus [pause] <u>co</u> came, *uh* the four boys in blue shirts *uh* didn't want to let the three small boys get into the bus first [pause]. After *uh* the four boys in blue shirts had been already on the bus, there were no seat left for the three small boys [pause]. As a result, the three small boys in red shirts [pause] had to wait for about half an hour for the next [pause] bus [pause]. And on three-thirty there was another bus [pause] <u>came</u> coming [pause]. <u>When the thr</u> when the three small boys got into the bus [pause], the bus rode to a countryside [pause]. And in the countryside *uh* the three small boys saw [pause] a broken bus [pause]. *Uh* this is the bus <u>which *uh* the four</u> [pause] which the four boys in blue shirts *uh* [pause] got into *uh* [pause] before them [pause]. *Um* as a result, the three small boys [pause] look at the four boys in blue shirt [pause], and [pause] they laugh at them [pause]. It is the end for the story.

Notes:

- 1. The underlined is where dysfluencies (e.g., repetitions, reformulations/selfcorrections, replacements, and false starts) were noted.
- 2. All pauses of 1 second or more were counted.
- 3. The italicized represents fillers.

### Writing Sample from the Planning-with-Writing Group

One day, 3 small boys in red shirts was were waiting at the bus stop. There are were some other people there. When the bus <u>number 26</u> arrived, the boys began getting on. Suddendly, there were 4 older boys who came after our 3 boys in red shirts jumped in <u>the middle of</u> the line. Therefore, the 3 boys had to <del>star</del> queue after those <del>rude</del> impolite young <del>man</del> men who were in blue shirts. Unfortunately, when the little boys in red shirts began getting on the bus, the <del>a</del> ticket taker told them that the bus <del>wa ca</del> could not amit <del>any</del> more passengers. It was 2 to 3 pm. <del>In th</del> Through the window of the bus <u>the</u> 4 boys in blue shirts laugh at them.

### Notes:

- 1. The crossed-out is where the participants themselves deleted words or phrases.
- 2. The underlined is where the participants themselves inserted new words or phrases.

# Appendix J: Inter-rater reliability

# Table 20: Inter-rater reliability

Categories	Measures	Ν	Pearson
	(frequency count)		correlation coefficient <i>r</i>
Fluency *	Total syllables	90	.96
	Filled pauses	90	.97
	Pauses end-clause	90	.99
	Pauses mid-clause	90	.99
	Repetitions	30	.98
	Self-reformulations	30	.97
	Replacements	30	.96
	False-starts	30	.94
Complexity **	T-units	90	.99
	Clauses	90	.99
	Verb variety	90	.93
	Total conjugated verbs	90	.99
	Type-token ratio	90	.99
	Content words	90	.98
Accuracy **	Total errors	90	.94
	Verb form errors	90	.96
	Article errors	90	.96
	Obligatory contexts for articles and inappropriate supplied articles	90	.96
	Plural noun ending errors	90	.97
	Obligatory contexts for plural noun endings and inappropriate supplied plural -s	90	.97
	Lexical errors	90	.98

*Note.* \* The measures were counted per original performance.

\*\* The measures were counted per pruned performance.

REFERENCES

#### REFERENCES

- Adams, R., & Ross-Feldman, L. (2008). Does writing influence learner attention to form? In D. Belcher & A. Hirvela (Eds.), *The oral/literate connection: Perspectives on L2 speaking, writing, and other media interaction* (pp. 210-225). Ann Arbor: University of Michigan Press.
- Anderson, J. (1995). Learning and memory: An integrated approach. New York: Wiley.
- Baddeley, A., & Logie, R. (1999). The multiple-component model. In A. Miyake & P. Shah (Eds), *Models of working memory*. New York NY: Academic Press.
- Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *Modern Language Journal*, *93*, 227-240.
- Bygate, M. (1996). Effects of task repetitions: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching*. Oxford: Heinemann.
- Bygate, M. (1999). Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks. *Language Teaching Research*, *3*, 185-214.
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2<sup>nd</sup> ed.). Boston, MA: Heinle & Heinle Publishers.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367-83.
- de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13, 1-24.
- Derwing, T. M., & Munro, M. J. (2001). What speaking rates do nonnative listeners prefer? *Applied Linguistics*, 22, 324–37.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359-380.
- Derwing, T. M., Munro, M. J., Thomson, R. E., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31, 533-557.
- Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R.
  Ellis (Ed.), *Planning and task performance in a second language* (pp. 219-238).
  Amsterdam: John Benjamins.

- Ellis, R. (1987). Interlanguage variability in narrative discourse: Style in the use of the past tense. *Studies in Second Language Acquisition*, *9*, 12-20.
- Ellis, R. (2001). Introduction: Investigation of form-focused instruction. *Language Learning*, *51*, 1-46.
- Ellis, R. (2003). Task-based language learning and teaching. Oxford University Press.
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics, 30*, 474-509.
- Ellis, R., & Barkhuizen, G. (2005). Analyzing learner language. Oxford University Press.
- Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167-192). Amsterdam: John Benjamins.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18,* 299-323.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354-375.
- Field, A. (2009). Discovering statistics using SPSS. California: SAGE Publications Inc.
- Gilabert, R. (2007). The simultaneous manipulation of task complexity along planning time and (+/-here-and-now): Effects on L2 oral production (pp.44-68). In M. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning*. Clevedon, UK: Multilingual Matters.
- Hardison, D.M. (2011, March). L2 engagement, perceived competence, and communication affect: Predictors of competence in the component skills of L2 oral interactions. Paper presented at the American Association for Applied Linguistics Conference, Chicago.
- Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, *11*, 329-350.
- Heaton, J. (1975). Beginning composition through pictures. Harlow: Longman.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*, 461-473.
- Ishikawa, T. (2007). The effect of manipulating task complexity along the +/- here-and-now dimension on L2 written narrative discourse. In M. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning*. Clevedon, UK: Multilingual Matters.

- Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of information processing approach to task design. *Language Learning*, 51, 401-436.
- Jackson, D., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A synthesis and metaanalysis of research on second language task complexity. *Language Learning*, 63, 330-367.
- Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 354-375). Cambridge, UK: Cambridge University Press.
- Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 143-164). Amsterdam: John Benjamins.
- Kim, Y. (2008). The effects of integrated language-based instruction in elementary ESL learning. *Modern Language Journal*, 92, 431-451.
- Kormos, J. (2006a). Fluency and automaticity in L2 speech production. *Speech production and second language acquisition* (pp.154-165). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kormos, J. (2006b). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P. Robinson (Ed.), Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance (pp. 39-60). Amsterdam: John Benjamins.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*, 145-164.
- Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, *17*, 48-60.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, *75*, 440-448.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307-322.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

- Levelt, W. J. M. (1999). Language production: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 83-122). Oxford: Oxford University Press.
- Liu, J., & Hansen, J. (2002). *Peer response in second language writing classrooms*. Ann Arbor: University of Michigan Press.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83-108.
- Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, *12*, 11-37.
- Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23, 451-468.
- Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*, 555-578.
- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21,* 108-148.
- Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77-109). Amsterdam: John Benjamins.
- Park, S. (2006). The influence of pretask instructions and pretask planning on learners' focus on form during task-based interaction. University of Illinois at Urbana-Champaign.
- Pathey-Chavez, G., & Ferris, D. (1997). Writing conferences and the weaving of multi-voiced texts in college composition. *Research in the Teaching of English*, *31*, 51-90.
- Payne, J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory and interlanguage development. *CALICO Journal*, 20, 7-32.
- Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, *6*, 69-79.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- Polio, C. (2001). Research methodology in second language writing research: The case of textbased studies. In T. Silva & P. K. Matsuda (Eds), On second language writing (pp. 91-115). Lawrence Erlbaum.

- Poole, A. (2005). The kinds of forms learners attend to during focus on form instruction: A description of an advanced ESL writing class. *The Asian EFL Journal*, 7(3), 75-87. Retrieved on June 14, 2011 from: <u>http://www.asian-efl-journal.com/September\_2005\_EBook\_editions.pdf</u>.
- Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36-57.
- Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437-470.
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes*, 14, 423-441.
- Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45, 99-140.
- Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge: Cambridge University Press.
- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27-57.
- Robinson, P. (2003). Attention and memory. In C. J. Doughty & M. H. Long (Eds), *The handbook of second language acquisition* (pp.631-678). Oxford: Blackwell.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in componential framework for second language task design. *International Review of Applied Linguistics*, 43, 1-32.
- Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In Garcia-Mayo, M.D.P. (Ed.), *Investigating tasks in formal language learning* (pp. 7-27). Clevedon, UK: Multilingual Matters.
- Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance (pp. 3-37). Amsterdam: John Benjamins.
- Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics*, 30, 533-554.

- Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111-141). Amsterdam: John Benjamins.
- Sasaki, T. (2004). Recipient orientation in verbal report protocols: Methodological issues in concurrent think aloud. *Hawaii Working Papers on Second Language Studies*, 22, 1-55.
- Segalowitz, N. (2010). Cognitive bases of second language fluency. New York: Routledge.
- Shirai, Y. (2002). The prototype hypothesis of tense-aspect acquisition in second language. In R. Salaberry & Y. Shirai (Eds), *The L2 acquisition of tense-aspect morphology*. Amsterdam: John Benjamins.
- Skehan, P. (1989). Individual differences in second language learning. Edward Arnold.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38-62.
- Skehan, P. (1998). A cognitive approach to language learning. Oxford: Oxford University Press.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*, 510-532.
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences of foreign language performance. *Language Teaching Research*, *1*, 185-211.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93-120.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge: Cambridge University Press.
- Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning* and task performance in a second language (pp. 193-216). Amsterdam: John Benjamins.
- Storch, N. (1999). Are two heads better than one? Pairwork and grammatical accuracy. *System*, 27, 363-374.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 471-484). Mahwah, NJ: Lawrence Erlbaum.
- Tarone, E., & Bigelow, M. (2005). Impact of literacy on oral language processing: implications for second language acquisition research. *Annual Review of Applied Linguistics*, 25, 77-97.

- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-273). Amsterdam: John Benjamins.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*, 84-119.
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.
- VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*, 287-301.
- Weissberg, R. (2000). Developmental relationships in the acquisition of English syntax: writing vs. speech. *Learning and Instruction*, *10*, 37-53.
- Weissberg, R. (2006). *Connecting speaking and writing in second language writing instruction*. Ann Arbor: University of Michigan Press.
- Wendel, J. (1997). Planning and second language narrative production. Unpublished doctoral dissertation, Temple University Japan, Tokyo.
- Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*, 85-106.
- Williams, J. (1999). Learner-generated attention to form. Language Learning, 49, 583-625.
- Williams, J. (2004). Tutoring and revision: Second language writers in the writing center. Journal of Second Language Writing, 13, 173-201.
- Williams, J. (2008). The speaking-writing connection in second language and academic literacy development. In D. Belcher & A. Hirvela (Eds.), *The oral/literate connection: Perspectives on L2 speaking, writing, and other media interaction* (pp. 10-25). Ann Arbor: University of Michigan Press.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. –Y. (1998). Second language development in writing: Measures of fluency, accuracy, and complexity. University of Hawai'i, Second Language Teaching and Curriculum Center.
- 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, 1-27.