The effects of planning on second language oral performance in Japanese: processes and production

Takako Nakakubo

University of Iowa

Copyright 2011 Takako Nakakubo

This dissertation is available at Iowa Research Online: http://ir.uiowa.edu/etd/1038

Recommended Citation


Follow this and additional works at: http://ir.uiowa.edu/etd

Part of the First and Second Language Acquisition Commons
THE EFFECTS OF PLANNING ON SECOND LANGUAGE ORAL PERFORMANCE
IN JAPANESE: PROCESSES AND PRODUCTION

by

Takako Nakakubo

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Second Language Acquisition
in the Graduate College of
The University of Iowa

May 2011

Thesis Supervisors: Professor Yukiko A. Hatasa
Associate Professor Judith E. Liskin-Gasparro
ABSTRACT

For over two decades, studies on task planning and its role in second language learners’ oral performance have shown that the opportunity to plan for a task generally improves learners’ speech (Ellis, 2005). It has been hypothesized that task planning reduces cognitive load during language processing, thus allowing learners to attend to various aspects of language, and that this enhanced attention, in turn, results in more successful task performance. However, one limitation to this task planning research to date is that most studies have examined the effects of planning before task performance and have largely ignored the effects of planning that occur during task performance (Yuan & Ellis, 2003). The best methods for investigating the effects of pre-task and on-line planning remain unclear as well. Another limitation in planning research is that findings have been based exclusively on external observation and measurement of learners’ oral production; we know little about what strategies learners use while engaging in planning that may result in higher-quality speech. The present study attempted to fill these gaps.

The participants in this study (intermediate and high-intermediate university learners of Japanese) were divided into experimental groups and performed a narrative task under one of four planning conditions: a) no planning, b) planning before speaking performance, c) planning during performance, and d) planning both before and during performance. As an underpinning investigation for the effects of pre-task and on-line planning, no planning and pre-task planning groups were further divided into two planning conditions: with and without time pressure during task performance. Participants received a set of six-line drawings that depicted a story and were asked to retell the story in Japanese. To examine the effects of planning on task performance, fluency, complexity, and accuracy in the participants’ speech were analyzed. For the analysis of planning strategies, retrospective interviews were given to a group of the participants from each planning group immediately after the task performance.
The results indicate that participants who performed the task under time pressure spoke significantly faster that those who had no time pressure during the task, which suggests that provision of time pressure may not be an appropriate method to limit on-line planning in order to more efficiently examine effects of pre-task planning. Regarding the effects of pre-task and on-line planning on learner’s task performance, no significant differences were found in participants’ oral production across planning conditions, except in the area of lexical complexity. Participants without a pre-task planning opportunity produced narrative stories with a greater variety of vocabulary than those who planned before the task. A trade-off effect between lexical complexity and accuracy was found when participants planned either before or during the task. Another trade-off effect was found between lexical complexity and fluency for the participants with on-line planning only. The analyses of strategy use showed that second language learners generally selected similar strategies regardless of planning conditions. These results provided important pedagogical implications and suggested useful future research directions.

Abstract Approved:

________________________________________
Thesis Supervisor

________________________________________
Title and Department

________________________________________
Date

________________________________________
Thesis Supervisor

________________________________________
Title and Department

________________________________________
Date
THE EFFECTS OF PLANNING ON SECOND LANGUAGE ORAL PERFORMANCE
IN JAPANESE: PROCESSES AND PRODUCTION

by
Takako Nakakubo

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Second Language Acquisition
in the Graduate College of
The University of Iowa

May 2011

Thesis Supervisors: Professor Yukiko A. Hatasa
Associate Professor Judith E. Liskin-Gasparro
Copyright by
TAKAKO NAKAKUBO
2011
All Rights Reserved
CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Takako Nakakubo

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Second Language Acquisition at the May 2011 graduation.

Thesis Committee: 

Yukiko A. Hatasa, Thesis Supervisor

Judith E. Liskin-Gasparro, Thesis Supervisor

Timothy Ansley

Michael Everson

Carol Severino
To my parents
ACKNOWLEDGMENTS

The work of writing a dissertation takes much time, effort, and commitment, and this dissertation was no exception. Along the way there have been many unexpected obstacles and setbacks. I have no doubt that this dissertation project would have never been completed without the countless people who supported me directly and indirectly. I sincerely believe that I am a most fortunate person to have many professors, colleagues, friends, and family members right beside me, always generously providing me with encouragement and support.

First and foremost, I would like to thank my committee members. My deepest thanks go to my dissertation supervisors, Professors Yukiko Hatasa and Judith Liskin-Gasparro. I am grateful to Professor Hatasa for her continuous support. Although she was in Japan, she encouraged me as much as though she were here in Iowa the whole time. Professor Liskin-Gasparro advised me from the very beginning of this dissertation study when it started as a class project for her course. They are outstanding teachers and exceptional mentors. They taught me what good research should be by guiding me through this research project.

I am also thankful for the rest of the committee. I would like to thank late Professor Kathy Heilenman for her warm, but practical advice on both professional and personal levels, which meant very much to me. Kathy, you are greatly missed. Professor Timothy Ansley assisted me in understanding the statistical analyses. I appreciate his great patience and also words of encouragement that he never failed in giving me whenever I visited him. Professor Michael Everson provided me with valuable suggestions from his in-depth knowledge of Asian languages and Asian language education. Professor Carol Severino kindly accepted to become a member of my committee in the middle of the project when Professor Heilenman’s untimely death
precluded her from serving on the committee. I am grateful to Professor Severino for her advice based on her extensive research on second language writing.

In addition to my committee members, I would like to thank Professor Roumyana Slabakova who was my adviser for the first years of my graduate study. I appreciate her guidance and friendship, which made a great difference in my life at graduate school.

I am indebted to many people who provided me with technical support to complete my dissertation. First of all, I would like to thank staff members at the Language Media Center at the University of Iowa for their support for my data collection and analyses: Director Sue Otto, Becky Bohde, Carolyn Goddard, and Greg Johnson. A number of people in the Japanese Program generously supported my project. Many thanks go to Akiko Hagiwara, Chie Muramatsu, Chiemi Hanzawa, Mayumi Ikeda, and Curt Warner for their help with transcribing and coding the data. I am also thankful to Nobuaki Takahashi who gave his time to discuss ideas and draw illustrations for the narrative tasks for this project.

For the statistical analysis, I received tremendous support from the Statistics Outreach Center at the University of Iowa. I am thankful for Professor Sheila Barron and graduate assistants Jin Gong and Anthony Fina. Melissa Fan also provided me with statistical help.

I am thankful to Roberta Golliher for her assistance in editing. She not only proofread the dissertation, but also provided me with kind words of encouragement.

There are many professors outside the University of Iowa who assisted me in various stages of this project. I would like to thank Professor Noriko Kobayashi at Tsukuba University, Japan, for allowing me to use SPOT for my project and kindly sending me all the necessary materials. Professor Lourdes Ortega at the University of Hawai‘i at Manoa directed me in finding one of her articles on planning. Professor Atsushi Fukada at Purdue University and Professor Masumi Tajima at Chuo Gakuin University, Japan, generously provided me with assistance in lexical analysis of my data.
I am obliged to the professors at participating institutions who kindly allowed me to recruit their students to participate in my study. Without their assistance and the participation of their students, this project would never have been possible.

I received funding from some organizations at the University of Iowa for my dissertation project. I am thankful to the Center for Asian and Pacific Studies and University of Iowa Student Government for their funding, which helped me travel to collect data. The Graduate College also provided me with generous funding, and I am most grateful for the support.

Last but not least, I would like to thank my family. During my graduate studies, I experienced great sorrow at my father’s passing. Looking back over my childhood, I have no hesitation to say that I learned everything that I needed to go through my dissertation project from my parents. They taught me the value of diligence, patience, and persistence and also the importance of education. But, above all I am most grateful that they always encouraged me to try with my best efforts even when the chance of success was slim. They taught me that everything is a learning experience and that I should just keep moving forward because I will learn something valuable no matter what the results are. And thus, I dedicate my dissertation to my parents with my deepest gratitude.

I would like to thank my in-laws, Barbara Wakeham and all the rest of the Wakehams, who have been understanding and supportive throughout my graduate studies. My husband, Brian, has been instrumental in my success through all phases of my dissertation. From editing to emotional support, he was there for me whenever I need help. Thank you for always believing that I could carry out this dissertation project even when I thought I could not. I thank my daughter, Nina, for keeping me grounded and reminding me of what is truly important. I am privileged and fortunate to have you both in my life.
ABSTRACT

For over two decades, studies on task planning and its role in second language learners’ oral performance have shown that the opportunity to plan for a task generally improves learners’ speech (Ellis, 2005). It has been hypothesized that task planning reduces cognitive load during language processing, thus allowing learners to attend to various aspects of language, and that this enhanced attention, in turn, results in more successful task performance. However, one limitation to this task planning research to date is that most studies have examined the effects of planning before task performance and have largely ignored the effects of planning that occur during task performance (Yuan & Ellis, 2003). The best methods for investigating the effects of pre-task and on-line planning remain unclear as well. Another limitation in planning research is that findings have been based exclusively on external observation and measurement of learners’ oral production; we know little about what strategies learners use while engaging in planning that may result in higher-quality speech. The present study attempted to fill these gaps.

The participants in this study (intermediate and high-intermediate university learners of Japanese) were divided into experimental groups and performed a narrative task under one of four planning conditions: a) no planning, b) planning before speaking performance, c) planning during performance, and d) planning both before and during performance. As an underpinning investigation for the effects of pre-task and on-line planning, no planning and pre-task planning groups were further divided into two planning conditions: with and without time pressure during task performance. Participants received a set of six-line drawings that depicted a story and were asked to retell the story in Japanese. To examine the effects of planning on task performance, fluency, complexity, and accuracy in the participants’ speech were analyzed. For the analysis of planning strategies, retrospective interviews were given to a group of the participants from each planning group immediately after the task performance.
The results indicate that participants who performed the task under time pressure spoke significantly faster than those who had no time pressure during the task, which suggests that provision of time pressure may not be an appropriate method to limit on-line planning in order to more efficiently examine effects of pre-task planning. Regarding the effects of pre-task and on-line planning on learner’s task performance, no significant differences were found in participants’ oral production across planning conditions, except in the area of lexical complexity. Participants without a pre-task planning opportunity produced narrative stories with a greater variety of vocabulary than those who planned before the task. A trade-off effect between lexical complexity and accuracy was found when participants planned either before or during the task. Another trade-off effect was found between lexical complexity and fluency for the participants with on-line planning only. The analyses of strategy use showed that second language learners generally selected similar strategies regardless of planning conditions. These results provided important pedagogical implications and suggested useful future research directions.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... xi

LIST OF FIGURES ........................................................................................................ xiv

CHAPTER 1: INTRODUCTION .................................................................................. 1
  1.1 Introduction ........................................................................................................... 1
  1.2 Research questions .............................................................................................. 4

CHAPTER 2 LITERATURE REVIEW ....................................................................... 7
  2.1 Introduction ........................................................................................................... 7
  2.2 Role of attention .................................................................................................. 7
  2.3 The output hypothesis ......................................................................................... 12
  2.4 Form-focused instruction .................................................................................... 15
  2.5 Classification of planning .................................................................................... 17
    2.5.1 Pre-task planning .......................................................................................... 17
      2.5.1.1 Rehearsal ................................................................................................ 18
      2.5.1.2 Strategic planning .................................................................................... 19
    2.5.2 On-line planning ........................................................................................... 20
    2.5.3 Other forms of planning ................................................................................ 22
  2.6 Previous studies on planning ................................................................................ 23
    2.6.1 Pre-task planning .......................................................................................... 23
      2.6.1.1 Rehearsal ................................................................................................ 23
      2.6.1.2 Strategic planning .................................................................................... 24
    2.6.2 On-line planning ........................................................................................... 27
    2.6.3 Trade-off effects ............................................................................................ 28
    2.6.4 Processes of planning .................................................................................... 30
  2.7 Limitations of the previous research ................................................................... 33
  2.8 Focus of the present study ................................................................................... 35

CHAPTER 3: METHOD ............................................................................................... 38
  3.1 Introduction ........................................................................................................... 38
  3.2 Participants ........................................................................................................... 38
  3.3 Materials ............................................................................................................... 43
    3.3.1 Questionnaire ................................................................................................ 44
    3.3.2 SPOT ............................................................................................................. 45
    3.3.3 Materials for the narrative tasks .................................................................... 47
    3.3.4 Materials for the retrospective interviews ..................................................... 49
  3.4 Task conditions .................................................................................................... 51
  3.5 Procedures ............................................................................................................ 56
  3.6 Measures .............................................................................................................. 61
    3.6.1 Measures for L2 oral production .................................................................... 61
      3.6.1.1 Data source: Transcriptions of oral production ........................................ 61
      3.6.1.2 Measures ................................................................................................ 63
        3.6.1.2.1 Fluency measures .............................................................................. 64
        3.6.1.2.2 Complexity measures ...................................................................... 67
        3.6.1.2.3 Accuracy measures .......................................................................... 69
      3.6.1.3 Interrater reliability .................................................................................. 74
    3.6.2 Evaluation of learner strategies ....................................................................... 75
      3.6.2.1 Strategy classifications .......................................................................... 75
      3.6.2.2 Interrater reliability to evaluate strategies ................................................. 84
CHAPTER 4: RESULTS OF L2 ORAL PRODUCTION ................................................. 88
4.1 Introduction ................................................................................................. 88
4.2 Selection of measures for speech data ......................................................... 88
4.3 Effects of task conditions ............................................................................ 96
  4.3.1 Time pressure ......................................................................................... 98
  4.3.2 Task planning ......................................................................................... 108
    4.3.2.1 Fluency .................................................................................. 114
    4.3.2.2 Complexity ............................................................................ 115
      4.3.2.2.1 Syntactic complexity .................................................... 115
      4.3.2.2.2 Lexical complexity ......................................................... 116
    4.3.2.3 Accuracy .................................................................................. 117
  4.3.2.4 Trade-off effects ............................................................................. 118
4.4 Summary of quantitative results ................................................................. 122

CHAPTER 5: RESULTS OF INTERVIEW ANALYSES ........................................... 124
5.1 Introduction .................................................................................................. 124
5.2 Effects of time pressure ............................................................................. 124
5.3 Strategy use ................................................................................................. 128
  5.3.1 Introduction .......................................................................................... 128
    5.3.1.1. Strategy use during pre-task planning .................................. 129
    5.3.1.2. Strategy use during task performance ......................... 131
    5.3.1.3. Comparison between pre-task planning and on-line planning ...... 134
  5.3.2 Selection of strategies by planning conditions ........................................ 136
    5.3.2.1. Metacognitive strategies ..................................................... 136
    5.3.2.2. Cognitive strategies .......................................................... 140
    5.3.2.3. Social/affective strategies ............................................... 142
  5.3.3 Depth of language analyses ................................................................. 143
5.4 Summary of the results of interview analyses ......................................... 146

CHAPTER 6 DISCUSSION .................................................................................... 148
6.1 Introduction .................................................................................................. 148
6.2 Discussion of Results ................................................................................ 149
  6.2.1 Discussion for research question 1: Effects of time pressure .................. 149
  6.2.2 Discussion for research question 2: Effects of planning ....................... 151
    6.2.2.1 The effects of task planning on the fluency of L2 oral production ...... 152
    6.2.2.2 The effects of task planning on the complexity of L2 oral production ........................................ 156
    6.2.2.3 The effects of task planning on the accuracy of L2 oral production ........................................ 162
    6.2.2.4 Trade-off effects among the fluency, complexity, and accuracy of L2 oral production ........................................ 166
  6.2.3 Discussion for research question 3: Strategy use .................................... 168
  6.2.4 Summary of discussion ........................................................................ 170
6.3 Pedagogical Implications ................................................................. 172
6.4 Limitations of the Present Study ..................................................... 174
6.5 Directions for Future Research ....................................................... 176
6.6 Conclusion ....................................................................................... 178

APPENDIX A: QUESTIONNAIRE .............................................................. 180

APPENDIX B: TASKS .......................................................... 181
Familiarization Task ............................................................................. 181
Main Task ............................................................................................. 182

APPENDIX C: TASK INSTRUCTIONS .................................................... 183

APPENDIX D: SAMPLE QUESTIONS IN THE SEMISTRUCTURED
RETROSPECTIVE INTERVIEW ................................................................. 186

APPENDIX E: STIMULATED RECALL INSTRUCTION ............................ 187

APPENDIX F: FACTOR ANALYSIS ......................................................... 188

APPENDIX G: CORRELATIONS AMONG MEASURES ............................ 189

REFERENCES ....................................................................................... 191
LIST OF TABLES

Table 1. Summary of student populations ................................................................. 41
Table 2. Participants’ native language ........................................................................ 42
Table 3. Task conditions ............................................................................................. 54
Table 4. Sample transcription conventions ................................................................. 62
Table 5. Measures for L2 oral production ................................................................. 64
Table 6. Particles in this study .................................................................................... 72
Table 7. Interrater reliability for 29 data samples ...................................................... 75
Table 8. Definitions and examples of metacognitive strategies .................................. 77
Table 9. Definitions and examples of cognitive strategies ........................................... 80
Table 10. Definitions and examples of social/affective strategies ............................... 83
Table 11. Experimental groups to be compared for research question 2 ................. 86
Table 12. Rotated component matrix ........................................................................... 89
Table 13. Correlations among fluency measures ....................................................... 92
Table 14. Correlations among complexity measures .................................................. 93
Table 15. Correlations among accuracy measures ..................................................... 94
Table 16. Descriptive statistics for participants’ narrative stories ............................... 98
Table 17. Task conditions for G1, G2, G3, and G4 ..................................................... 99
Table 18. Descriptive statistics for time in seconds that was spent to complete the task ................................................................................................................. 100
Table 19. Descriptive statistics for speech rate ........................................................... 101
Table 20. Descriptive statistics for the number of clauses per T-unit ....................... 102
Table 21. Descriptive statistics for type-token ratio .................................................... 102
Table 22. Descriptive statistics for percentage of error-free clauses ....................... 103
Table 23. Descriptive statistics for percentage of correct use of particles .............. 103
Table 24. Wilks’ lambda for a two-way MANCOVA for time pressure and pre-task planning effects ................................................................. 104
Table 25. Follow-up ANCOVA for time pressure and pre-task planning effects on speech rate ................................................................. 105

Table 26. Follow-up ANCOVA for time pressure and pre-task planning effects on the number of clauses per T-unit ........................................... 106

Table 27. Follow-up ANCOVA for time pressure and pre-task planning effects on TTR .................................................................................. 106

Table 28. Follow-up ANCOVA for time pressure and pre-task planning effects on the percentage of error-free clauses .............................................. 107

Table 29. Follow-up ANCOVA for time pressure and pre-task planning effects on the percentage of correct use of particles ........................................ 107

Table 30. Task conditions for G1, G3, G5, and G6 ................................................................. 108

Table 31. Descriptive statistics for speech rate ........................................................................ 109

Table 32. Descriptive statistics for the number of clauses per T-unit .................................. 110

Table 33. Descriptive statistics for type-token ratio ............................................................... 111

Table 34. Descriptive statistics for percentage of error-free clauses ...................................... 112

Table 35. Descriptive statistics for percentage of correct use of particles ............................... 112

Table 36. Wilks’ lambda for a two-way MANCOVA for pre-task planning and online planning effects on independent variables .................................. 113

Table 37. Follow-up ANCOVA for task planning conditions on speech rate ....................... 114

Table 38. Follow-up ANCOVA for task planning conditions on the number of clauses per T-unit ........................................................................ 115

Table 39. Follow-up ANCOVA for task planning conditions on TTR .................................. 116

Table 40. Follow-up ANCOVA for task planning conditions on the percentage of error-free clauses .................................................................... 117

Table 41. Follow-up ANCOVA for task planning conditions on the percentage of correct use of particles .................................................................. 118

Table 42. Correlations among dependent variables for G1, G3, G5, and G6 (n = 93) ... 119

Table 43. Correlations among dependent variables for G1 (n = 24) ........................................ 120

Table 44. Correlations among dependent variables for G3 (n = 24) ........................................ 120

Table 45. Correlations among dependent variables for G5 (n = 24) ........................................ 121

Table 46. Correlations among dependent variables for G6 (n = 21) ........................................ 122

Table 47. Participants’ perceptions of time pressure (n = 8) .................................................... 125
Table 48. Strategy use by the Pre-Task Planning group before the task......................... 130
Table 49. Strategy use by the No On-line Planning group during task performance ..... 132
Table 50. Strategy use by the On-line Planning group during task performance......... 133
Table 51. Participants’ use of metacognitive strategies by group............................ 137
Table 52. Participants’ use of cognitive strategies by group................................. 141
Table 53. Participants’ use of social/affective strategies by group.......................... 143
Table 54. Descriptive statistics for the number of moras by group....................... 159
Table 55. Descriptive statistics for the number of meaningful moras by group......... 159
Table 56. Wilks’ lambda for a two-way MANCOVA for pre-task planning and on-line planning effects................................................................. 160
Table 57. Follow-up ANCOVA for task planning conditions on moras................ 161
Table 58. Follow-up ANCOVA for task planning conditions on meaningful moras..... 161
LIST OF FIGURES

Figure 1. Procedures .............................................. 60
CHAPTER 1: INTRODUCTION

1.1 Introduction

For more than two decades, researchers in second language acquisition (SLA) have explored the effects of the opportunity to plan for a task on adult language learners’ oral performance. The line of studies on the interaction between planning and oral performance in a second language (L2) was initiated by Ellis (1987). These studies are generally grounded in information processing theory, “which claims that humans possess a limited processing capacity and, as a result, are not able to attend fully to all aspects of a task” (Yuan & Ellis, 2003, p. 1). Although it is not clear whether the capacity of cognitive processing is limited or unlimited (DeKeyser, Salaberry, Robinson, & Harrington, 2002; Robinson 2003; VanPatten, 2002a, 2002b), it has been generally agreed that there is a certain extent of constraint in the allocation of attentional resources while processing language (Baddeley, 2003; Baddeley & Logie, 1999; Pashler, 1998; Robinson, 2003; Skehan, 1998). L2 learners, due to their limited language proficiency, have to decide to which aspect of language to allocate their attention (Robinson, 2003; Skehan, 1998; VanPatten, 2002a, 2002b; Yuan & Ellis, 2003), because they are unable to focus, cognitively speaking, on all of the aspects of a speaking task at the same time. To better understand the interaction between cognitive load and language production, SLA researchers have turned to task planning and have examined its role in oral production in a second language. It is hypothesized that task planning reduces cognitive load during language production processing and allows the L2 learners to attend to various aspects of language to retrieve information in working memory, which is expected to result in a more successful task performance (Ellis, 2005).

Another theoretical motivation for task planning research comes from Swain’s output hypothesis (Swain, 1985, 1995, 2005). As a reaction to SLA models that claim that receiving input is sufficient for language acquisition (Krashen, 1985), Swain (1985)
developed the Output Hypothesis, which posits the importance of output for interlanguage development (Swain, 1985, 1995, 2005; Swain & Lapkin, 1995). With the opportunity to plan for a task, learners, it is generally suggested, produce better output (Bygate, 2001; Crookes, 1989; Ellis, 1987; Foster & Skehan 1996; Gilabert, 2007; Kawauchi, 2005a, 2005b; Mehnert, 1998; Ortega, 1995a, 1999, 2005; Sangarun, 2005; Skehan & Foster, 1997, 2005; Tajima, 2003; Wendel, 1997; Wigglesworth, 1997; Yuan & Ellis, 2003). Therefore, task planning has the potential to facilitate language learning.

The notion of focus on form also offers a rationale for investigating the role of task planning in oral production (Ellis, 2005; Ortega, 1999). Focus on form refers to the processing that directs learners’ attention to linguistic forms when learners are involved in communicative language use and is claimed to be a necessary condition for interlanguage development (Doughty & Williams, 1998a). Ortega (1995a, 1999, 2005) has reported that learner-initiated focus on form has occurred during the planning phase, which suggests positive effects of task planning on L2 learning.

In planning studies, there are two primary areas of research: (a) which aspect of language benefits the most when L2 learners have an opportunity to plan for a task, and (b) what L2 learners specifically do to ease the limitation of attention to produce better output. With respect to the first area, previous studies have investigated the effects of planning on fluency, complexity, and accuracy of L2 learners’ oral production (Bygate, 2001; Crookes, 1989; Elder & Iwashita, 2005; Ellis, 1987, 2005; Ellis & Yuan, 2005; Foster & Skehan 1996; Gilabert, 2007; Kawauchi, 2005a, 2005b; Mehnert, 1998; Mochizuki & Ortega, 2008; Ortega, 1995a, 1995b, 1999, 2005; Sangarun, 2005; Skehan & Foster, 1997, 2005; Tajima, 2003; Tavokoli & Skehan, 2005; Wendel, 1997; Wigglesworth, 1997; Wigglesworth & Elder, 2010; Yuan & Ellis, 2003). There are two types of planning: pre-task planning and on-line planning. For pre-task planning, learners receive time to plan for a task prior to the main performance, whereas on-line planning occurs during the task performance. The majority of past studies on planning have
explored the effects of pre-task planning and have generally found positive effects on fluency and complexity in L2 oral performance. However, its effect on accuracy has not been conclusive (Ellis, 2005, 2009; Ortega, 1999). The results of some studies revealed more accurate production by pre-task planners (Crookes, 1989; Sangarun, 2005; Tavokoli & Skehan, 2005), whereas other studies obtained results that show partial or no effects of pre-task planning on accuracy (Elder & Iwashita, 2005; Ellis, 1987; Foster & Skehan, 1996; Gilabert, 2007; Kawauchi, 2005a, 2005b; Mehnert, 1998; Mochizuki & Ortega, 2008; Ortega, 1995a, 1999; Skehan & Foster, 1997, 2005; Tajima, 2003; Wendel, 1997; Wigglesworth, 1997; Wigglesworth & Elder, 2010; Yuan & Ellis, 2003).

Regarding on-line planning, thus far only a few studies have examined the effects of this planning type (Ellis & Yuan, 2005; Hulstijn & Hulstijn, 1984; Skehan & Foster, 2005; Yuan & Ellis, 2003), and these studies have obtained mixed results. Yuan and Ellis (2003) have reported positive effects of on-line planning on grammatical complexity and accuracy; however the speed of speech may have decreased. Hulstijn and Hustijin (1984) and Skehan and Foster (2005) have found no clear effects of on-line planning on L2 learners’ speech. One of the problems in the studies for on-line planning is that it is difficult to determine whether or not learners are in fact engaged in planning while performing a task. The studies by Hulstijn and Hustijin (1984) and by Ellis and Yuan (Ellis & Yuan, 2005; Yuan & Ellis, 2003) have attempted to encourage learners to plan on-line by providing unlimited time to perform a task, while providing time pressure for the group of learners without on-line planning opportunities. Skehan and Foster (2005) have taken a different approach. They have provided new information in the middle of the decision-making task to force learners to plan on-line. More research exclusively on the effects of on-line planning as well as on the approaches to trigger on-line planning is necessary to determine its role in L2 oral production.

In addition to the effects of planning on fluency, complexity, and accuracy in L2 learners’ speech individually, the results of some studies have suggested that there are
trade-off effects; that is, one aspect improves while another suffers because limited attention is available. These trade-off effects have been observed between fluency and accuracy, between complexity and accuracy, and between grammatical accuracy and the variety of vocabulary (Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003).

With respect to research on what learners specifically do to ease the limitation of attention to produce better output, a limited number of studies have been conducted to date (Kawauchi, 2005a; Ortega, 1995a, 1999, 2005; Sangarun, 2005). These studies have shown that L2 learners actively attend to grammatical form during planning, although the degree of focus on grammatical form depends on whether their primary focus is on the accuracy of their oral production or on communication (Ortega, 1995a, 1999, 2005; Sangarun, 2005). Also, the results of Kawauchi’s (2005a) study have revealed that language learners at different levels of proficiency use different strategies during planning. Although these studies have provided some insight into learners’ cognitive processes, our understanding of what and how L2 learners plan for a task is far from complete. It is necessary to further research the processes of planning to identify what L2 learners do while engaging in planning, whether pre-task or on-line, to understand the effects of planning on L2 oral performance.

1.2 Research questions

An accumulation of the findings in the task planning studies suggests that planning improves second language (L2) learners’ production (Ellis, 2005; Ortega, 2005). Yet, significant questions still remain to be answered. One limitation is that most studies have examined the effects of planning that takes place before task performance, rather than during (or before and during) the performance of the task (Yuan & Ellis, 2003). In addition, there is an issue of inconsistency in measurement across past planning studies (Ellis, 2005, 2009; Ortega, 1999); that is, past planning studies have used different measures to evaluate L2 oral production, which makes comparison of findings among
studies difficult. Another methodological limitation is that findings have been based exclusively on external observation and measurement of learners’ oral production; we have little knowledge about what choices L2 learners make and what strategies they use while planning that may result in higher-quality speech (Ortega, 2005). Finally, as is the case in other areas of SLA research, the L2 performance investigated in most of the planning studies to date is English or another Western language. Research on less commonly taught languages, such as Japanese, is needed to obtain a universal picture of the effects of planning across languages. The current study attempts to fill these gaps in the L2 planning research.

There are three main research questions in this study regarding the effects of task planning on L2 learners’ oral production.

1. How does time pressure affect L2 learners’ oral production in a narrative task?
2. Are there differences in the way pre-task and on-line planning affect L2 learners’ oral production in a narrative task?
   2.1 How do the different types of planning affect the fluency of L2 learners’ oral production in a narrative task?
   2.2 How do the different types of planning affect the complexity of L2 learners’ oral production in a narrative task?
   2.3 How do the different types of planning affect the accuracy of L2 learners’ oral production in a narrative task?
   2.4 Are there any trade-off effects among the fluency, complexity, and accuracy of L2 learners’ oral production in a narrative task?
3. What do L2 learners do during planning?
Research question 1 addresses the effects of time pressure. On-line planning is involved to some degree in all speech; when investigating pre-task planning, results may be affected according to the amount of on-line planning that L2 learners engage in during the task. Yuan and Ellis (2003) have attempted to minimize the effects of on-line planning by providing a time limit for the task performance when examining the effects of pre-task planning. However, it is possible that learners speak faster because of the time pressure, not as a result of pre-task planning. Therefore, it is necessary to determine whether or not the time pressure affects L2 learners’ fluency. Research question 2 explores the effects of different types of planning to determine how each type of planning interacts with L2 learners’ oral production. Research question 3 attempts to reveal the overall processes of task planning and to discover the strategies L2 learners use during planning.

In the next chapter, the literature for attention and its role in language acquisition is reviewed. The literature on two theoretical frameworks in SLA that are relevant to the research on planning, Swain’s output hypothesis (Swain, 1985, 1995, 2005) and focus on form, is also reviewed. Finally, the classification of planning types and the review of past studies on each type of planning are presented. In chapter 3, the methodology employed in this study is described. Chapter 4 presents the results of the analyses of the speech data and chapter 5 reports the results of learners’ strategy use to plan for a task. Chapter 6 discusses the implications of the results and presents suggestions for future research.
CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter first reviews the role of attention, which is one of the important factors that affect learners’ production. Then it discusses two theoretical rationales for task planning research: the output hypothesis and focus on form. The chapter continues with the description of types of planning and illustrates how each type of planning has been carried out in previous studies on task planning, followed by findings of past task planning studies. Finally, the research questions are presented.

2.2 Role of attention

According to Baddeley and Logie (1999), working memory consists of a phonological loop, visuo-spatial sketch pad, and central executive component. The phonological loop processes acoustic information, whereas the visuo-spatial sketch pad processes graphic information. These two components exist independently, but both of them are controlled by the central executive which, in turn, is controlled by attention.

The question of how learners process language is a central issue in the field of SLA. Particularly, the concept of attention and its role in language acquisition is one of the most extensively researched and discussed topics to date. Although there has been a problem of vague definitions of attention and other similar terms such as consciousness and awareness (Tomlin & Villa, 1994), it has been generally agreed that attention is essential for L2 acquisition to occur (Long, 1996). Tomlin and Villa (1994) review the research of cognitive processes in SLA as well as in cognitive science, and they analyze conceptions of attention in depth. According to their fine-grained analysis, attention comprises three functions: alertness, orientation, and detection. Tomlin and Villa (1994) also explain that awareness is “a particular state of mind in which an individual has undergone a specific subjective experience of some cognition content or external stimulus” (p. 193). In their view, awareness is different from consciousness because
consciousness includes multiple meanings, such as perception, understanding, and intention. This study follows Tomlin and Villa’s (1994) definitions of the terms attention, awareness, and consciousness unless specified otherwise.

In his ongoing research, Schmidt (1990, 1995, 2001) has argued that attention plays a central role in L2 acquisition. Referring to research in psychology and cognitive science, Schmidt states that “attention is necessary for all aspects of L2 learning” (2001, p. 3). As basic assumptions on attention in psychology, Schmidt (2001) presents the following six points (pp. 11-16):

1. attention is limited;
2. attention is selective;
3. attention is subject to voluntary control;
4. attention controls access to consciousness;
5. attention is essential for the control of action; and
6. attention is essential for learning.

These six assumptions are intertwined with each other, but assumptions 1, 3, and 6 in the list above are directly related to aspects of task planning. First, according to Schmidt (2001), the principal view of attention in psychology is that it involves a limited capacity for cognitive processing. Strong support for this concept comes from studies in working memory. Baddeley and Logie (1999) have proposed that the central executive, a main component of working memory, plays a vital role in controlling attention. Based on empirical studies conducted in neurological science, Baddeley and Logie (1999) suggest that the phonological loop and the visuo-spatial sketch pad, the two specialized systems in working memory, have capacity constraints. These systems serve the central executive, which results in limiting the capacity of the central executive (Baddeley, 2003; Baddeley & Logie, 1999).
Research in psychology also suggests capacity limits for attention. In reviewing the literature on psychological experiments on attention, Pashler (1998) found consistent results that could be interpreted as capacity overload. For example, when research participants are involved in processing multiple stimuli simultaneously, their performances become less accurate as the stimuli become more complex. Although there may be considerable differences between processing language for learning and non-language stimuli, these findings indicate the possibility of limitations in attentional capacity.

In SLA research, several researchers support the proposal that attention is constrained by limited capacity for human information processing (Schmidt, 2001; Skehan, 1998; Skehan & Foster, 2001; VanPatten, 1990, 2002a. 2002b, 2004). VanPatten (1990) conducted a study that compared learners’ listening comprehension when they were asked to pay attention to meaning only or to both meaning and form simultaneously. The results revealed that the learners, especially those with lower L2 proficiency, had difficulty attending to form while processing meaning, suggesting that meaning and form compete for attentional resources. However, DeKeyser, Salaberry, Robinson, and Harrington (2002) criticized VanPatten (2002a), asserting that current theories of attention in psychology consider attention to be unlimited. In response, VanPatten (2002b) argued that their claim was based on models of attention for first language (L1) speakers, not L2 speakers. Given that L2 acquisition is more complex than L1 acquisition, involving multiple factors, VanPatten (2002b) claimed that L1 processing models are not applicable to L2 processing.

Robinson (2003) acknowledges that there are restrictions on the ability to process linguistic information, but questions the idea that capacity limitation is the cause for various processing problems. He claims that it is not capacity limits, but a limit in the control of cognitive functions and interference during attentional resource allocation that are responsible for processing overload (Robinson, 2003).
Although there still exists much discussion and controversy regarding capacity limitation in attention (whether it is limited or unlimited), research in various disciplines including SLA has revealed that attention undergoes certain constraints during processing depending on task difficulty and cognitive demands (Baddley, 2003; Baddley & Logie, 1999; Pashler, 1998; Robinson, 2003; Skehan, 1998). Regardless of the cause of attentional constraints, the opportunity for task planning is considered to reduce the attention load during the performance (Ellis, 2005). When planning time is available, learners are able to allocate their attention to various aspects of language to retrieve information in working memory before starting to perform the task; thus, it is possible that more proficient speech will be produced.

Schmidt’s (2001) third notion of attention is its voluntary nature. Schmidt (2001) states that preparatory attention has been found to be beneficial for both comprehension and production in terms of accuracy and other factors, and although SLA researchers have not examined this type of attention yet, it is linked with studies of planning (Hulstijn & Hulstijn, 1984; Ortega, 1999). When provided with time to plan, learners have an opportunity to decide what to pay attention to, which can be meaning, form, or both meaning and form. The decision can be made consciously or unconsciously and also can be initiated by the learners or guided by the task instructions (Crookes, 1989; Foster & Skehan, 1996; Hulstijn & Hulstijn 1984; Ortega, 1999; Skehan & Foster, 2005; Sangarun, 2005; Wendel, 1997). Sangarun (2005) investigated the effects of guided planning. The participants in her study were instructed to plan for a task under one of three conditions: focusing on meaning, focusing on form, or focusing on both meaning and form. Sangarun (2005) reported that the participants were generally able to allocate their attention as instructed. Ortega (1999) conducted retrospective interviews with the participants in her study on pre-task planning and she found that the participants intentionally attended to form during planning even though they received no instruction to do so. The degree of focus on form depended on whether their primary focus was on accuracy of their oral
production or on the communication of content. Some of the participants were more attuned toward accuracy and, therefore, attended more to grammar, while others were more concerned with communication, which lead their attention to the content of their utterance. Schmidt (2001) states that people have the ability to attend to one specific stimulus with or without an opportunity to plan, but these studies on planning show that learners can voluntarily control their attention to certain aspect of language regardless of the presence of a specific instruction to do so.

Finally, Schmidt (2001) claims that attention is necessary for learning. He explains that any learning, including language learning, requires the stimuli to be stored in long-term memory, and attention is essential for this process. In SLA research, VanPatten (2002a, 2002b, 2004) has also argued that learners need to attend to input for further processing. Leow (1998) conducted a study to investigate the necessity of attention, specifically the necessity of detection, following Tomlin and Villa’s (1994) term, for L2 learning. The results showed that participants who detected novel grammatical items were able to recognize these forms more frequently and were able to use the forms more accurately than those who did not detect them. These results suggest that detection leads to learning at least at a morphological level.

Most of the past studies on attention in SLA have concerned its role in input processing (Schmidt 2001; VanPatten 1990, 2002a, 2004), specifically converting input into intake. However, the role of attention is not limited to input and intake, but it also plays a critical part in output. Swain (2005) claims that “the activity of producing the target language may prompt second language learners to recognize consciously some of their linguistic problems” (p. 474). Attention is needed when producing output to retrieve the linguistic knowledge learners have, and planning is considered to lessen the stress of this process by creating extra capacity in attention resources. If learners are able to allocate their attention to forms during planning as studies have suggested, it can be assumed that planning also contributes to learning.
2.3 The output hypothesis

The output hypothesis was first presented by Swain (1985). The basic idea of this hypothesis is that language production, whether speaking or writing, plays an active role in facilitating L2 learning when certain conditions are met (Swain, 2005). In the 1980s, the mainstream premise in SLA research was based on input as represented by Krashen’s input hypothesis (Krashen, 1985). Krashen (1985) states that language acquisition takes place when learners receive input that is slightly above their current proficiency level (i+1). In his position, language production emerges as a consequence of acquisition, but it has no contribution to language acquisition.

Swain (1985) questioned Krashen’s claims based on her observations of French immersion classes. Although immersion students exhibited higher proficiency in listening and reading compared to students in regular French as a foreign language programs, speaking and writing abilities were not developed to a degree similar to that of comprehension. Swain (1985, 2005) pointed out that what was lacking for students in the immersion program was production, particularly production that is pushed to be appropriate both grammatically and sociolinguistically.

In a series of articles (Swain, 1995, 1998, 2005), Swain presents three functions of production in the processes of second language learning: 1) the noticing/triggering function, 2) the hypothesis-testing function, and 3) the metalinguistic function. The noticing/triggering function of output is that learners may notice something about their interlanguage, specifically that they lack the linguistic knowledge to express what they want to communicate, as they try to speak or write in the L2 (Swain, 1995, 1998, 2005). This noticing leads learners to realize that they do not know a certain linguistic form. Swain (1998, 2005) suggests that there are multiple levels of noticing. For example, learners may notice a certain L2 form because it appears frequently or because it has salient features. Learners may also notice that there are some differences between the L2 form in discourse produced by native speakers and the form in their interlanguage.
Finally, learners may notice that they do not know how to deliver their message in the L2. Swain (1995, 1998, 2005) claims that noticing may prompt learners to generate new L2 knowledge or to strengthen the knowledge that they have already gained.

The second function of output is hypothesis testing (Swain, 1995, 1998, 2005). Learners may examine their hypothesis of how some aspects of the L2 works by trying it out in speaking or writing. In other words, learners may produce language to test whether it is comprehensible and whether it is formed correctly (Swain, 1995). By doing so, learners are expanding the current level of their interlanguage. At times, learners receive feedback on their output, which may trigger reformulation of their hypothesis. Swain and Lapkin (1995) propose that the process of this type of reformulation is one of the ways that language learning takes place.

Last, Swain (1995, 1998, 2005) discusses the metalinguistic function of output; that is, speaking serves as a mediating tool to reflect on language production. Learners may talk about someone else’s or their own language production, and by talking about language, what they notice and the hypotheses they form and test will become accessible for examination (Swain, 1998). Output enables learners to articulate, transform, and reflect on their thoughts and to draw their attention to inconsistencies, if there are any.

Referring to linguistic output within a sociocultural theoretical framework, Swain (2005) states its role as follows:

Speaking (and writing) are conceived of as cognitive tools—tools that mediate internalization; and that externalize internal psychological activity, resocializing, and recognizing it for the individual; tools that construct and deconstruct knowledge; and tools that regulate and are regulated by human agency. (p. 480)

Swain (2005) states that even research methods such as think-aloud and stimulated-recall protocols serve processes of learning, and should not be considered as data collection tools only. Think-alouds and stimulated recalls assist learners in understanding and restructuring their experiences, and they play an important role in learning.
De Bot (1996) reviews Swain’s output hypothesis from a psycholinguistic perspective and explains that one of the important roles of producing output is to transform learners’ declarative knowledge into procedural knowledge. To describe the development of procedural knowledge, de Bot (1996) states that language production involves various stages of form-function connection.

When the output, as a result of a certain connection, does not elicit negative feedback, the connection will be reinforced. As the connection repeatedly recurs, the processes of the connection become faster and more specific, and eventually all of the related rules and factors of the connection will be consolidated into learners’ receptive knowledge. When learners become aware that output does not correspond with their receptive knowledge, the form-function connection will be hindered from forming. This awareness may be triggered internally by examining the output through their knowledge or externally by receiving negative feedback from their interlocutors. De Bot (1996) states that this is how noticing, one of the functions of output that Swain (1995, 1998, 2005) suggests, occurs.

Although the interlocutor of the learner plays an important role, benefits of output for language acquisition are not fully dependent on external feedback (de Bot, 1996; Swain, 2005; Swain & Lapkin, 1995). For example, Swain and Lapkin (1995) state that noticing a gap in their linguistic knowledge may occur spontaneously when learners have trouble producing output, and external feedback is not always necessary to trigger noticing. Also, with regard to hypothesis testing, de Bot (1996) explains that learners develop internal speech and monitor its form and meaning before they speak. Since learners’ receptive knowledge is generally more consistent than their productive knowledge, they are able to reform and improve their production as they examine their hypotheses internally. When learners produce an utterance after going through these processes, their output reaches its optimal level if the system of internal hypothesis testing is working properly. When learners have an opportunity to plan, they can afford to
spend time to evaluate their internal speech, and it is possible that they notice a gap in their L2 knowledge and test their hypothesis against their internal norm. In this sense, planning can assist learners in producing better output, which may lead to learning.

In addition, de Bot (1996) argues that noticing a gap and trying to solve the problem of inconsistencies between the message and linguistic form by themselves, rather than by receiving the correct form in the input, results in more successful learning. When learners detect a problem and actively search their memory for the solution, they focus on a particular aspect of production processes. As they get involved in analyzing the source of the problem, their metalinguistic knowledge is strengthened. The deeper the analysis, the more solid the memory becomes, which results in interlanguage development.

2.4 Form-focused instruction

*Focus on form* has been discussed as the rationale for task planning research in recent years (Ellis, 2005; Ortega, 1999). It is a pedagogical approach that is generally presented in contrast with *focus on forms* and *focus on meaning*. What distinguishes the notion of focus on form from those of the other two terms is the way that learners process linguistic and communicative information. Doughty and Williams (1998a) summarize the fundamental difference in the three concepts by explaining that “focus on form entails a focus on formal elements of language, whereas focus on forms is limited to such a focus, and focus on meaning excludes it” (Doughty & Williams, 1998a, p. 4, emphasis in original). Focus on form is a condition where learners’ attention is drawn precisely to a linguistic feature as necessitated by a communicative demand (Doughty & Williams, 1998a). Reviewing the definitions of focus on form that have been proposed in the past research, Doughty (2001) states that “focus on form involves learners briefly and perhaps simultaneously attending to form, meaning, and use during one cognitive event”
(Doughty, 2001, p. 211). Doughty and Williams (1998b) claims that focus on form is more beneficial for learning than focus on forms or focus on meaning.

Reviewing several studies of focus on form (Lightbown & Spada, 1990; Lyster, 1994; Spada & Lightbown, 1993; White, 1991), Long and Robinson (1998) reported generally positive effects of focus on form on learning. However, they criticized that many of the studies assumed the occurrence of focus on form by simply providing instruction that was rich in focus on form. The level of attention that learners allocated to form and meaning cannot be scrutinized unless researchers look into learners’ conscious mental processes upon receiving form-focused instruction. It is also helpful to examine what learners are doing internally because, due to the nature of classroom dynamics, it is difficult to find out what aspects of instruction directly resulted in learning (Lyster, 1994). Long and Robinson (1998) suggested the use of finely tuned measures, such as questionnaires, to examine to what extent the instructional treatments actually triggered the intended focus on form.

Ellis (2005) states that focus on form is one of the fundamental constructs for language processing that underlies theoretical frameworks of the study of planning. The concept of focus on form has been explored in three contexts: pedagogic, discoursal, and psycholinguistic contexts (Ellis, 2005). Among them, Ellis (2005) states that focus on form from a psycholinguistic perspective is one of the central constructs to consider in theoretical frameworks for planning research. He explains that the opportunity to plan for a task extends the limits of attentional capacity and allows learners to shift their attention to formal aspects of the language while preparing for meaningful language use, which may enable them to connect form and meaning. Because of the time allowed for planning, learners are able to access their knowledge, even knowledge that is not available for automatic processing (Ellis, 2005). By examining retrospective interviews, Ortega (1995a, 1999, 2005) reported that learner-driven focus on form occurred during task planning in the two studies she conducted in 1995 and 1999.
The preceding overview of the role of attention, output hypothesis, and focus on form provides some support that task planning may create optimal conditions for language learning to take place. During cognitive processing, attention undergoes certain constraints. Since L2 learners have limited language proficiency, they are not able to allocate their attention to all of the aspects of oral production. The opportunity to plan for a task may reduce the cognitive load that is involved in speaking by creating extra capacity in attention resources. From the perspective of output hypothesis, task planning may assist L2 learners in producing better output by allowing them to spend time to evaluate their linguistic knowledge. Furthermore, when learners are allowed to plan for a task, they may be able to shift their attention to form while being engaged in meaningful language use, which may lead to language learning.

Turning to task planning, there are different types of planning. In the next section, each type of planning that has been investigated in prior studies is described in detail.

2.5 Classification of planning

As principal types of planning, Ellis (2005) presents pre-task planning and within-task planning. The latter is also referred to as on-line planning (Yuan & Ellis, 2003). The major difference between the two is the timing of the planning with respect to task performance; that is, whether the planning occurs before or during the task performance. The following sections deal with each type of planning, including illustrative examples from research on task-based planning.

2.5.1 Pre-task planning

Pre-task planning takes place prior to task performance. Ellis (2005) points out that pre-task planning is different from pre-task activities, such as brainstorming, in terms of access to the task materials. During pre-task planning, learners receive the actual materials for the task, whereas during the pre-task activities, learners do not have access to the materials that they are going to use to perform the task.
There are two types of pre-task planning: rehearsal and strategic planning. Although both rehearsal and strategic planning require that learners be engaged in some activities to prepare for the task, there is a distinct difference between them. The next section describes rehearsal, followed by the explanation of strategic planning.

2.5.1.1 Rehearsal

Ellis (2005) defines rehearsal as task repetition; that is, learners perform “the same or slightly altered tasks—whether whole tasks, or parts of a task” (Bygate & Samuda, 2005, p. 43). In Bygate (2001), a group of participants watched a short cartoon video with no dialogue and told the story that the video described. Ten weeks later, the participants performed exactly the same task. They watched the same cartoon video that they had watched before and again told the story that was described in the cartoon. The learners received no feedback on the task or on their task performance during the ten-week period.

The interval of time between the initial performance and the repeated performance varies across studies. For example, in the study by Bygate (2001), the interval was ten weeks, while in Kawauchi’s (2005a, 2005b) study, the initial and repeated performances took place within a short time period. The participants in her study performed a narrative task, responded to a questionnaire, and then performed the same task for a second time (Kawauchi, 2005a, 2005b). In addition, the participants performed the same task for a third time one week later (Kawauchi, 2005a).

When repeated performance improves over the first performance, it is considered that learning has occurred between the two task performances. However, the degree that the first performance has influenced the repeated performance in terms of learning may depend on the length of the interval. When there is a long interval between the two performances, the improvement that was observed in the repeated performance may suggest that the memory from the first performance remains effective for a long time. At
the same time, improved performance may be the result of learning unrelated to the effects of rehearsal. On the other hand, if the interval between the first and the second performances is short, the improvement in the second performance may be considered as the outcome of repeating the task. However, it cannot be determined if there is any long-term effect on learning of repeating a task.

2.5.1.2 Strategic planning

The second type of pre-task planning is strategic planning. For this type of pre-task planning, learners receive a period of time prior to task performance, during which they deliberate on the information they need to deliver and how to convey it to carry out the task. In other words, strategic planning gets learners involved in thinking about the content and the language they need for the task performance. Although pre-task planning includes both rehearsal and strategic planning in precise meaning, many studies use the term, pre-task planning, to indicate only strategic planning (Crookes, 1989; Foster & Skehan, 1996; Kawauchi, 2005a, 2005b; Mehnert, 1998; Ortega, 1999, 2005; Skehan & Foster, 1997, 2005; Wendel, 1997; Wigglesworth, 1997; Yuan & Ellis, 2003).

The planning time in the previous studies varies from one minute (Mehnert, 1998; Wigglesworth, 1997) to one hour (Ellis, 1987), but in the majority of the studies the learners had a ten-minute preparation time (Crookes, 1989; Foster & Skehan, 1996; Kawauchi, 2005a, 2005b; Mehnert, 1998; Ortega, 1999; Skehan & Foster, 1997, 2005; Wendel, 1997; Yuan & Ellis, 2003).

In most of the cases, learners were allowed to take notes to prepare for the main performance (Crookes, 1989; Foster & Skehan, 1996; Kawauchi, 2005a, 2005b; Ortega, 1999; Sangarun, 2005; Skehan & Foster, 2005; Wendel, 1997; Yuan & Ellis, 2003), but there are a variety of other activities for strategic planning, such as writing a composition, reading a story, and practice. For example, instead of taking notes, the participants in the studies by Ellis (1987) and Kawauchi (2005a, 2005b) wrote a composition about the story
that they later told as the task. Kawauchi (2005a, 2005b) also investigated the role of reading and practice as strategic planning. For reading, one group of her participants received a model story describing a set of pictures that they were going to narrate in the main task. They were instructed to think how they could tell the story using the reading as a reference. Another group of participants in Kawauchi’s (2005a, 2005b) study were told to practice the task by saying aloud what they were going to say when telling the story.

2.5.2 On-line planning

On-line planning, which is also referred to as within-task planning, is a type of planning that is available while engaging in task performance. Yuan and Ellis (2003) propose the definition of on-line planning as follows:

On-line planning is the process by which speakers attend carefully to the formulation stage during speech planning and engage in pre-production and post-production monitoring of their speech acts. (p. 6)

Ellis (2005) classifies on-line planning into two categories: pressured and unpressured planning. The difference between them is the time given to learners for the task performance. In pressured on-line planning, learners have a time limit to complete the task, which restricts their time to engage in planning during the task. On the other hand, in unpressured planning, learners are allowed to spend as much time as they wish on the task. By providing the learners with unlimited time, researchers assume that learners will engage in on-line planning while performing the task.

Skehan and Foster (2005), however, pointed out that it is simply an assumption that the learners would engage in on-line planning if they have no time restriction during the task performance. In the absence of data on learner behavior, there is no way to know whether the learners are planning their utterances on-line. In Skehan and Foster (2005), the learners performed a decision-making task. To make sure that the learners planned
while performing the task, Skehan and Foster (2005) provided new information so that the participants would be forced to alter their decision and to change what they were going to say.

Although pre-task planning and on-line planning contrast with each other, it is possible to employ both types of planning for task performance. Ellis (2005) suggests four possible planning conditions: (a) no pre-task or within-task planning, (b) pre-task planning, but with no within-task planning, (c) no pre-task planning, but with within-task planning, and (d) both pre-task and within-task planning. Under condition (a), when the learners are allowed neither pre-task nor on-line planning, they perform the task immediately after receiving the task instruction. Under condition (b), the learners have the opportunity to plan for the task before the performance, but have limited time during the task, thus minimizing on-line planning. Under condition (c), in contrast, the learners have no opportunity to plan before the task performance, but they have unlimited time to carry out the task. Under condition (d), the learners receive time both before and during the task so that they are able to engage in both pre-task and on-line planning.

Most of the past studies (Crookes 1989; Foster & Skehan, 1996; Kawauchi, 2005a, 2005b; Mehnert, 1998; Ortega, 1999; Skehan & Foster, 1997, 2005; Wendel, 1997; Yuan & Ellis, 2003) focused on the effects of planning by comparing student performance under conditions (a) and (b). However, only a few studies (Skehan & Foster, 2005; Yuan & Ellis, 2003) have investigated the impact of condition (c) on task performance, and hardly any studies have been done to date on condition (d). Therefore, our understanding of how planning interacts with L2 learners’ oral production is limited to when they have an opportunity to plan before the task, and we have little knowledge of how planning affect L2 learners’ task performance when they plan during the task, or when they plan both before and during the task. To understand the role of planning in L2 performance as a whole, more research on on-line planning and the combination of pre-task and on-line planning is necessary.
2.5.3 Other forms of planning

In addition to classifying planning types by when the planning takes place, there are various forms of planning that are different in structure (Ellis, 2005). One of those varieties is guided and unguided planning, which is also referred to as detailed and undetailed planning respectively (Foster & Skehan, 1996; Skehan & Foster, 2005). Under the guided planning condition, the learners receive specific instructions on what and how they should prepare for the task, whereas under the unguided planning condition, learners receive no such instruction and plan on their own for the task. These task conditions have been applied to both pre-task and on-line planning studies (Foster & Skehan, 1996; Hulstijn & Hulstijn, 1984; Sangarun, 2005; Skehan & Foster, 2005; Wendel, 1997).

The instructions for guided planning vary in terms of the degree of specification and in the areas where the learners should focus their attention. Sangarun (2005) provided a specific series of steps for learners to follow to form their arguments during strategic planning. On the contrary, Wendel (1997) provided general suggestions to instruct the learners to attend to vocabulary and the story line when planning. Regarding the focus of the instructions, learners may be asked to pay attention to the grammar, vocabulary, content, or structure of their oral production. Those instructions can be classified into three types: focus on meaning, on form, or on both meaning and form (Crookes, 1989; Foster & Skehan, 1996; Sangarun 2005; Skehan & Foster, 2005; Wendel, 1997).

In summary, there are two distinct types of planning: pre-task planning and on-line planning. Each type can be further divided into two types: rehearsal and strategic planning for the pre-task planning, and pressured and unpressured planning for on-line planning. Besides the difference in allocation of time to prepare for a task, planning can be categorized in terms of differences in its structure. These categories of planning include guided and unguided planning. Past research has explored all of the planning types either individually or combined and has investigated the effects on learners’ task performance.
2.6 Previous studies on planning

Several studies have been conducted on the effects of task planning, particularly strategic planning, on L2 oral performance. This section reviews findings of those studies on each type of task planning. Most of them have compared the fluency, accuracy, and complexity of L2 learners’ oral production under planning and no-planning conditions. In addition to the effects of task planning on those three aspects, research findings on trade-off effects as well as analyses of planning processes are reviewed.

2.6.1 Pre-task planning

2.6.1.1 Rehearsal

Bygate (2001) explored the effects of rehearsal on fluency, accuracy, and complexity on L2 learners’ oral performance for narrative and interview tasks. The participants in his study performed the same task twice with an interval of ten weeks between the two performances. Although the data showed that the second performance was characterized by greater fluency and complexity than the first performance, task repetition did not result in improved accuracy. These results generally correspond to those in Kawauchi (2005a). In her study, the participants performed the same task three times: an unplanned performance, a planned performance (with pre-task planning activities), and a delayed performance. Comparing the unplanned and delayed performances, the researcher found that fluency improved in the delayed performance at a statistically significant level for both low-intermediate and advanced learners. Similarly, overall complexity was higher in the delayed performance by low-intermediate learners, although only one measure, the number of subordinate clauses, reached significance, but not the number of clauses per T-unit or the number of words per T-unit. Regarding advanced learners’ performance, a significant difference was found only in the number of clauses per T-unit, indicating that the effect of task repetition on complexity is limited for both low-intermediate and advanced learners, but in the different ways. Low-intermediate
learners were significantly more accurate in terms of the use of past tense markers, whereas no significant difference was found for advanced learners.

Ko (2001) assigned the participants in her study to perform a story-telling task twice under one of two task conditions: with or without a session for negotiation of meaning between the first and second performances. She compared how the participants improved the quality of their tellings and found that the second performance under both task conditions generally showed some improvement. Although the focus of Ko’s (2001) study was not specifically the effects of rehearsal on L2 learners’ speech, the results provided support for the contention that learners improve their speech simply by repeating a task.

Using randomly selected data from Bygate (2001), Bygate and Samuda (2005) investigated discourse complexity by analyzing framing in learner speech. Framing refers to any descriptions in a story additional to a basic narrative, as well as the perspectives of the speaker, listener, or characters in the story. They found that the amount of framing increased significantly in the second task over the first task. Bygate and Samuda (2005) discuss that by repeating the same task, learners may gain more attentional capacity while performing the task for the second time since the content of the story is better organized compared to the first performance, which allows learners to produce a more complex discourse structure. This study suggests that the effects of rehearsal on L2 learners’ oral performance are not limited to fluency, accuracy, and syntactic complexity, and that a holistic approach to analyzing learners’ speech may be necessary to uncover the benefits of task planning.

2.6.1.2 Strategic planning

Strategic planning has been most extensively researched among three types of planning. One of the influential studies was conducted by Yuan and Ellis (2003). They compared learners’ oral narrative performance under no planning and strategic planning
conditions, and they obtained results that generally supported the findings of previous studies (Crookes 1989; Foster and Skehan, 1996; Mehnert, 1998; Ortega, 1995a, 1999; Wendel, 1997). Yuan and Ellis used two measures for fluency: the number of syllables per minute and the number of meaningful syllables per minute. To calculate meaningful syllables, they excluded repeated, reformulated, or replaced parts in the speech were excluded. Yuan and Ellis (2003) found that the planning group produced more syllables per minute in both measures than the no planning group did. Although these differences between the two groups failed to reach statistical significance, the results of other studies such as Ortega (1999), Tajima (2003), and Wendel (1997) revealed that the speech rate under the planned condition was significantly higher than that under the unplanned condition.

Greater fluency for planners based on other types of fluency measures, breakdown fluency and repair fluency (Skehan & Foster, 2005; Tavakoli & Skehan, 2005) was also found. For example, Skehan and Foster (1997) reported that the number of pauses (1 second or longer) was larger for non-planners than for planners with a 10-minute pre-task planning opportunity. Kawauchi (2005b) calculated the proportion of repetitions in participants’ oral production to examine fluency. The participants in her study were intermediate, high-intermediate, and advanced ESL learners. Kawauchi (2005b) found that intermediate and high-intermediate learners who planned prior to the task performance repeated themselves less than non-planners at the same proficiency levels did. However, advanced learners were negatively affected by planning and produced more repetition when they were able to plan.

Turning to complexity, the strategic planning group in Yuan and Ellis (2003) produced syntactically more complex speech than the no planning group with a statistically significant difference, which again replicated the results of previous studies (Crookes 1989; Foster and Skehan, 1996; Mehnert, 1998; Ortega, 1999; Wendel, 1997). The complexity measures used in those studies generally included the number of words
per utterance, the number of clauses per T-unit, and the number of subordinate clauses per T-unit. Kawauchi (2005b) reported that the participants in her study produced more complex speeches in terms of the number of words per T-unit and the number of clauses per T-unit regardless of their L2 proficiency level. However, for the frequency of subordinate clauses per two-minute performance, there was no difference between planners and non-planners among advanced learners, whereas planners at intermediate and high-intermediate levels used more subordinate clauses than their non-planning peers.

Previous planning studies have not examined lexical complexity as extensively as syntactic complexity, but some studies investigated the effects of planning on the vocabulary of the subsequent L2 oral production. Most of the planning studies that investigated lexical complexity used either type-token ratio (Ortega, 1995a, 1999), or measures based on type-token ratio (Gilabert, 2007; Tajima, 2003; Yuan & Ellis, 2003). The strategic planning group in the study of Yuan and Ellis (2003) outperformed the no planning group, but the result did not reach statistical significance. Ellis (2009) states that pre-task planning may positively affect syntactic complexity rather than lexical complexity.

Kawauchi (2005a, 2005b) examined the effects of strategic planning on lexical density, which is a ratio of lexical words such as nouns, verbs, and adjectives, to the total number of words produced by the learners. She found that participants across three proficiency levels showed significant improvement in lexical density when they had an opportunity to plan.

Although prior studies generally suggested a positive impact of planning on the fluency and complexity of L2 learners’ oral production, mixed results were obtained regarding accuracy. The results of Yuan and Ellis (2003) revealed that the strategic planning group was more accurate than the no planning group when comparing error-free clauses and correct verb forms, but not at a statistically significant level. Crookes (1989) also found that the planning group performed more accurately than the no-planning group,
but the difference was not significant. The results of Mehnert (1998) reached a statistically significant level between non-planners and planners. Concerning these contradictory findings, Mehnert (1998) claimed that different measures for accuracy contributed to different results. Some studies used specific measures such as plural forms (Crookes, 1989; Wigglesworth, 1997) and articles (Ortega, 1999), whereas others used more general measures such as percentage of error-free clauses (Foster & Skehan, 1996; Wigglesworth & Elder, 2010; Yuan & Ellis, 2003).

Another possible explanation for the mixed results for accuracy is learner proficiency level. Kawauchi (2005a, 2005b) employed percentage of correct use of past tense to measure accuracy in oral production by intermediate, high-intermediate, and advanced learners. The results revealed that intermediate and high-intermediate learners performed more accurately at a statistically significant level when they had an opportunity to plan. However, there was no difference between planners and non-planners among the advanced-level learners in her study.

In sum, research on strategic planning has generally found positive effects of pre-task planning on fluency and complexity in L2 speech. However, its effect on accuracy has not been conclusively shown because past studies have obtained mixed results. Yuan and Ellis (2003) pointed out that one of the reasons for these mixed results could be planning while the learners were engaged in the task. That is, those studies did not consider how much time the participants spent on the task and, therefore, there was a possibility that a different amount of on-line planning was allowed depending on the study. Further research is necessary to investigate exclusively the effects of strategic planning by controlling the amount of on-line planning.

2.6.2 On-line planning

Turning to the effects of on-line planning, a limited number of studies have examined this type of planning. First, Yuan and Ellis (2003) investigated the effects of
on-line planning on fluency, complexity, and accuracy, in addition to the effects of no planning and strategic planning. They reported that the no-planning group performed more fluently than the on-line planning group. Although there was no statistically significant difference between the two groups, this result indicates that on-line planning may decrease the speed of speech. As for complexity, statistically significant effects of on-line planning were found for grammatical complexity. Also, there was a statistically significant difference between the no-planning and on-line planning groups in terms of accuracy, the on-line planning group being more accurate than the no-planning group.

Skehan and Foster (2005) also investigated the effects of on-line planning in comparison with the effects of strategic planning. To make sure that the learners plan while performing a decision-making task, they provided new information that affects the learners’ decision in the middle of the task. The learners in their study worked in pairs and discussed what would be an appropriate sentence for people who committed various crimes, either killing or seriously injuring another person. For example, one of the cases was that of a woman who killed her husband after discovering that he had an affair with another woman. Five minutes into the task, new evidence about the crime was introduced, which in this case was that the husband had been violent and abusive to his wife and children and also had had a series of affairs. By providing the information that was not included in the original description of each case, Skehan and Foster (2005) attempted to force the learners to get involved in on-line planning. The performances during the first five minutes and the second five minutes were compared, but no clear effects of on-line planning were found. Skehan and Foster (2005) state that further research is necessary to determine the effects of on-line planning.

2.6.3 Trade-off effects

Due to the constraints in attentional capacity during task performance, it is hypothesized that different aspects of language compete for limited attentional resources
during processing (Ellis, 2005; Skehan 1998). As a result, one aspect of performance is prioritized and improves, whereas another aspect receives less attention and remains the same, if it does not get worse. Such phenomena are called trade-off effects (Skehan & Foster, 1997; Yuan & Ellis, 2003). Trade-off effects have been observed in various studies on task planning. Foster and Skehan (1996) reported one example of a trade-off effect between complexity and accuracy. The results of their study revealed that when the participants performed the Personal Information Exchange task, their speech production was the most accurate among the three tasks, while their complexity was the lowest. In contrast, the participants exhibited the highest complexity and the least accuracy in the Narrative task. Foster and Skehan (1996) concluded that limited information processing could not allow L2 learners to allocate their attention to both complexity and accuracy and, therefore, their performance resulted in either great complexity with poor accuracy or vice versa. Results of this study are echoed by Skehan and Foster (1997).

Mehnert (1998) also reported results that indicated a trade-off between complexity and accuracy. She compared the effects of different lengths of planning time and found that the participants prioritized accuracy when they had one minute of planning time, and that they gave a higher priority to complexity when they had ten minutes of planning time.

Based on their own results, Yuan and Ellis (2003) suggested that there are two types of trade-off effects for L2 learners with lower proficiency, due to limited cognitive processing capacity. The first trade-off is between fluency and accuracy. When the learners are planning on-line and are paying attention to accuracy, their fluency is negatively affected. If they are able to plan ahead, their fluency increases, but their accuracy decreases. The second trade-off is between grammatical accuracy and lexical variety. When learners attend to grammatical accuracy, their vocabulary becomes less varied, and when their vocabulary becomes rich, their accuracy goes down. Yuan and
Ellis (2003) proposed the possibility that there would be fewer trade-off effects if the L2 learners had enough time to plan both before and during the task.

The last three sections reviewed past planning studies. The researchers in most of these studies were primarily interested in the products (i.e., the language produced in the performance of the task) as a result of engaging in task planning. However, as the research on task planning progresses, the interest in the processes of planning, in addition to the products, has been increasing. In recent years, more studies have examined what L2 learners do while preparing for a task, using questionnaires (Kawauchi 2005a; Tajima 2003), retrospective interviews (Kawauchi 2005a; Ortega 1995a, 1999, 2005; Wendel, 1997), and think-aloud protocols (Park, 2006; Sangaru, 2005). The findings of these studies will be discussed in the following section.

2.6.4 Processes of planning

In the last 20 years, the majority of studies on task planning have been concerned with L2 learners’ oral production. We have learned a great deal about how the opportunity to plan before or during a task may improve some aspects of L2 speech. On the other hand, we have little knowledge about what L2 learners actually do to plan for a task. What choices do L2 learners make and what strategies do they use while planning that result in higher-quality speech?

Ortega (1995a, 1999) conducted one of the first studies that attempted to answer these questions, and she reported her findings on learners’ conscious use of strategies during pre-task planning based on retrospective interviews. The participants in her study performed story-retelling tasks three times: once for practice, once preceded by strategic planning, and once without planning. Each performance was immediately followed by a retrospective interview, using the researcher’s observations, her field notes and participants’ notes during planning, and the pictures for the task as stimuli to elicit accurate responses from the learners.
Ortega (1995a, 1999) presented the following strategies as the most commonly used approaches by the learners:

1. Sort out essential from inessential, known from unknown.
2. Identify what you don’t know, then proceed.
3. First work on main ideas, organization, or both, then work on details.
4. Don’t take unnecessary risks, that is, if they are not essential for the task.
5. Concentrate on identified problem spots.
6. When rehearsing, go in order first, then be flexible.
7. When practicing, go from more to less support, and go through the story in different ways.
8. Estimate your time before deciding what to do next.

(Ortega, 1999, p. 127)

Ortega (1995a, 1999) explained that the above strategies were neither an exhaustive list, nor were they presented in hierarchical order. Her findings clearly suggest that learners are inclined to use some kind of problem-solving strategies when engaged in planning for a task.

Ortega (1995a) also identified three types of cognitive strategies that most of the learners used while planning: writing, rehearsal, and memorization. First, the participants were allowed to write notes during the pre-task planning time, and 27 of the 32 participants took advantage of this opportunity (Ortega, 1995a). Half of the 27 reported that their activities during the entire ten minutes of planning involved their notes, whether writing, reading, or revising them, whereas the other half engaged in writing only during the first half of the planning time and spent the rest of the time on other activities, with or without their notes. The five informants who did not write notes stated that writing would not help them since they knew that they would not be allowed to look at their notes while performing the task. Second, a majority of the participants in Ortega’s (1995a) study rehearsed, mostly in their heads, although the manner of their rehearsal varied. Some participants engaged in a complete rehearsal, focusing on problematic parts of the story, while others practiced only the problematic areas. Third, regarding memorization, most
of the participants stated that one of the benefits of writing was better recall of vocabulary and grammar (Ortega, 1995a). They claimed that visualizing what they intended to say helped them remember when they performed the task.

One unique aspect of Ortega’s (1995a, 1999) study is that the participants were in dyads and that the listener had to put pictures in the correct order while the speaker was telling a story. Because of this feature of the task, Ortega (1995a, 1999) reports that the participants considered the impact of their performance on their partners and more likely attended to meaning so that their partners could complete the task appropriately.

Ten years after her first attempt to illustrate what learners did during the strategic planning phase, Ortega (2005) reanalyzed the same interview data from Ortega (1995a, 1999) as well as the data she obtained from the participants in Ortega (1995b). She transcribed the interviews and classified the learners’ strategies into three categories: metacognitive strategies, cognitive strategies, and social/affective strategies. Ortega (2005) found that the learners used metacognitive and cognitive strategies at almost the same frequency. Also, she found that advanced-level learners and low-intermediate learners used different types of strategies. Advanced learners put their efforts equally on retrieval and rehearsal strategies to prepare for task performance. They also used a variety of monitoring strategies, such as auditory monitoring and cross-language monitoring. These strategies were not used by many of the low-intermediate learners. The number of strategy types that learners used was also different between low-intermediate and advanced learners. Overall, advanced learners used a greater variety of strategies than low-intermediate learners.

Kawauchi (2005a) also compared strategy use during pre-task planning by low-intermediate and advanced learners. She reported that one of the major differences between these two groups was the focus of their planning. Low-intermediate learners attended more to the planning of vocabulary, and they tried to avoid ideas and reduce the content when they had trouble finding appropriate vocabulary to express themselves. On
the other hand, advanced learners were more concerned with the organization and content of stories and tried to find effective ways to convey their intended meanings.

Sangarun (2005) also explored L2 learners’ behaviors during planning by using plan-aloud protocols. As a result, she found that learners generally focused on meaning regardless of the type of planning instructions they received. On the other hand, there were some differences in what they did while planning, depending on the instructions.

2.7 Limitations of the previous research

A number of studies have been conducted for the past two decades to explore the effects of task planning on L2 learners’ performance. In summary, previous studies have shown positive effects of pre-task planning, particularly strategic planning, on L2 learners’ oral performance in terms of fluency and complexity, while the effects on accuracy are still unclear. Regarding the effects of on-line planning, Yuan and Ellis (2003) found positive effects on complexity and accuracy, but Skehan and Foster (2005) did not find any effects of on-line planning on L2 learners’ oral performance. To explain these inconclusive results and to better understand the interaction between task planning and L2 performance, Ortega (1999) has suggested a coherent research program. It would involve more consistent research designs and cumulative interlanguage measures, so that new results are comparable to those in the previous studies. It also involves a broader scope that allows researchers to thoroughly investigate different variables, such as task complexity and learner factors (Ortega, 1999).

The second limitation of task planning studies is that only a few studies (Skehan & Foster, 2005; Yuan & Ellis, 2003) have specifically investigated the effects of on-line planning. These studies used different operationalizations of on-line planning and obtained different results. Yuan and Ellis (2003) allowed the learners in the on-line planning group to spend unlimited time to perform a task so that they would engage in on-line planning, while the pre-task planning group had a time limit. One of their findings
is that the pre-task planning group spoke more fluently than the on-line group. However, it is not clear whether the pre-task planning group spoke faster because of the time pressure, not as a result of pre-task planning. Skehan and Foster (2005) attempted to trigger on-line planning by providing new information in the middle of task performance, and the results showed no effects of the treatment. More research exclusively on the effects of on-line planning is necessary to establish the operationalization of on-line planning and to investigate the interaction between on-line planning and L2 oral production.

Third, a variety of measures of fluency, complexity, and accuracy have been adopted in previous planning studies and, therefore, there has been less coherence in the results across these studies (Ellis, 2005; Ortega, 1999). The results of one study should be comparable with those of previous studies. Reliability and validity of measures are also an issue. There has been much discussion of whether each measure effectively assesses as it is supposed to do (Norris & Ortega, 2009).

Another point of concern for task planning research is to investigate the processes involved in internal planning. The majority of the previous studies have assumed that learners attend to language during planning by simply examining their output, and only a few of them have attempted to shed light on what was actually happening to learners internally (Ellis, 2005; Ortega, 1999; Skehan & Foster, 2005). Ortega (1995a, 1999, 2005) included retrospective interviews in her study, and she found that the participants actively paid attention to form. Insight into learners’ cognitive processes is rich with information for research on planning as Ortega (1995a, 1999, 2005) shows. Gass and Mackey (2007) mentioned Mehnert’s (1998) pre-task planning study as a prime example of one that would have benefited from retrospective interviews to explain what was going on in learners’ mind and what aspects of their thinking affected task performance. Future research should be directed to explore learners’ internal processes during planning and how they affect oral performance.
2.8 Focus of the present study

The present study attempts to fill the gaps that still remain in task planning studies. It explores three main research questions regarding the effects of task planning on L2 learners’ oral production. These questions are:

1. How does time pressure affect L2 learners’ oral production in a narrative task?

2. Are there differences in the way pre-task and on-line planning affect L2 learners’ oral production in a narrative task?
   2.1 How do the different types of planning affect the fluency of L2 learners’ oral production in a narrative task?
   2.2 How do the different types of planning affect the complexity of L2 learners’ oral production in a narrative task?
   2.3 How do the different types of planning affect the accuracy of L2 learners’ oral production in a narrative task?
   2.4 Are there any trade-off effects among the fluency, complexity, and accuracy of L2 learners’ oral production in a narrative task?

3. What do L2 learners do during planning?

Research question 1 addresses the effects of pre-task planning as well as the operationalization of on-line planning. On-line planning is involved to some degree in all speech and, therefore, Yuan and Ellis (2003) suggested controlling the effects of on-line planning with the time allowed to perform the task. They provided a time limit for the strategic planning group to minimize the on-line planning and provided no time limit for the on-line planning group to encourage planning as they talked. As a result, the fluency
of the strategic planning group improved while the fluency of the on-line planning decreased. However, it is possible that learners spoke faster because of the time pressure, not as a result of pre-task planning. At the same time, it is not clear whether or not learners actually plan on-line when they have no time limit to perform a task. Thus, it is necessary to determine how the time pressure affects L2 learners’ oral production.

Research question 2 investigates how different types of task planning affect the fluency, complexity, and accuracy of L2 learners’ oral production. Previous studies have found that L2 learners generally improve their task performance in terms of the fluency and complexity of their speech when they have an opportunity to plan prior to the task. However, the effects of on-line planning on L2 learners’ task performance is largely unknown to date, and also, we know very little about how planning interacts with L2 learners’ oral production when they are allowed to plan both before and during the task performance.

Another point that should be investigated is trade-off effects. Some planning studies have reported trade-off effects between complexity and accuracy (Foster & Skehan, 1996, Mehnert, 1998), between fluency and accuracy (Yuan & Ellis, 2003), and between grammatical accuracy and lexical variety (Yuan & Ellis, 2003) in L2 learners’ production. More research is necessary to identify how fluency, complexity, and accuracy interact, and how each type of planning affects the trade-off effects.

To answer research question 2, learner production should be assessed using the most reliable and valid measures. This study performed a factor analysis on some measures that were commonly used in previous planning studies to evaluate fluency, complexity, and accuracy in L2 oral production and attempted to select appropriate measures based on the results.

Research question 3 attempts to reveal what strategies L2 learners use to plan for the task. The majority of the past studies investigated primarily the outcome of planning,
but not the processes of task planning. This study focuses on what and how L2 learners plan for the task in addition to examining their oral production.

This chapter has reviewed the literature on attention. The output hypothesis and focus on form are also discussed. Then, the chapter has reviewed the findings of past task planning studies and has presented three research questions. The next chapter describes the populations in this study and the data collection procedure. The measures and the analyses that were used for this study are also explained.
CHAPTER 3: METHOD

3.1 Introduction

This chapter presents the methodology for the study, which consists of narrative tasks and retrospective interviews. It first describes the participants, followed by an explanation of the materials, design, and procedures. Finally, the measures and the analyses of L2 learners’ oral production and strategy use are explained.

3.2 Participants

The participants in this study were intermediate or high-intermediate learners of Japanese at nine university-level institutions. Pseudonyms are used for these nine participating schools: Schools A, B, C, D, E, F, G, H, and I. These schools were selected to collect data because they had relatively large Japanese programs and had at least 40 students at the target proficiency levels.

Schools A–F are large public universities in the Midwest with undergraduate enrollments of 20,000 to 36,000. Besides Schools B and F, the other four schools complete the elementary level textbooks in four semesters of instruction. School B completes the elementary level textbooks in two semesters, and School F does so in three semesters.

All of the participants from Schools G, H, and I were enrolled in intensive summer courses in Japanese when the data were collected. School G is a private college in New England. One of the unique aspects of School G is that it is a total immersion program where all the students pledge at the beginning of the course to use only the L2 as a means of communication both inside and outside of the classroom. The students at School G lived in dormitories during the course with their teachers, who were native speakers of Japanese. Schools H and I are academic institutions in Japan that are affiliated with American universities. The students at School H arranged their accommodation by themselves, and most of them lived in apartments, whereas the
students at School I stayed with Japanese host families during the course. The majority of
the participants at Schools G, H, and I had studied Japanese at universities in the United
States right before attending the summer programs. At the beginning of the intensive
courses, the students at these institutions took proficiency tests and were enrolled in the
appropriate course levels in Japanese.

As described above, the participants in this study were recruited from nine
institutions in three different settings: domestic regular courses (Schools A–F), an
immersion program (School G), and study abroad programs (Schools H and I). Thus,
learning context is one of the confounding variables in this study because it affects the
quantity and quality of L2 learners’ exposure to the target language, which may result in
differences in development of L2 abilities, including oral proficiencies (Collentine, 2004;
Collentine & Freed, 2004; Freed, Segalowitz, & Dewey, 2004; Segalowitz & Freed,
2004). It is possible that the participants’ task performance and strategy use while
engaging in the task may be different depending on the setting in which they were
studying Japanese, which would interfere with the results of this study.

To control the effects of learning settings, several points were taken into
consideration upon data collection. First, with few exceptions, all of the participants had
completed the elementary level of Japanese in a regular classroom setting at either in high
school or in a university in the United States. Also, most of the learners at School G
(immersion setting) and Schools H and I (study abroad setting) had studied Japanese in
domestic formal classrooms before being enrolling in the programs at the respective
schools. Therefore, their learning experiences of Japanese before the summer intensive
programs were comparable to those of the learners at Schools A–F. Second, to keep the
effects of different learning settings to the minimum, the data at Schools G, H and I were
collected at an earlier stage of each program, approximately one week to one month after
the summer courses began.
To further reduce the differences in quality and quantity of exposure to the Japanese language that originated from learning settings other than regular classroom, immersion, and study abroad contexts, potential participants who had a native speaker in their immediate family and those who lived in Japan when young were excluded from the study. Also, learners at Schools A–F who were living with a native Japanese speaker at the time of data collection and reported speaking with that person in Japanese regularly were excluded from this study, whereas those at Schools G–I in a similar circumstance were not. The language environment for the learners at Schools A–F is expected to remain relatively unchanged long term because they attend these schools as full-time regular students. Therefore, the learners at Schools A–F who had Japanese roommates might have had considerably more contact with the Japanese language in everyday life than the other learners for an extended time, which may have affected the results of this study. In contrast, the learners at Schools G–I were enrolled in summer courses. Their language environment, therefore, was a temporary condition and was expected to have minor effects on the results of this study.

A total of 147 Japanese learners from Schools A–I participated in this study. Out of 147 participants, the data from two participants were discarded from the sample due to recording problems that occurred during the data collection. The data from two other participants were also excluded since they performed the narrative task much differently from the rest of the participants; one participant performed the task as if it were a skit, and the other included several critiques on the task while telling the story. The data from the remaining 143 participants were analyzed to explore the effects of planning on their oral production. This study included retrospective interviews, and 18 participants from School A and six participants from School B participated in this portion of the study in addition to performing narrative tasks. Table 3.1 displays the participants’ demographic information for each student group, which was collected by a questionnaire described in the next section.
Table 1. Summary of student populations

<table>
<thead>
<tr>
<th>School</th>
<th>Number (Male/Female)</th>
<th>Mean age (Range)</th>
<th>Mean length of Japanese study (Range)</th>
<th>Mean length of stay in Japan (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>36 (21/15)</td>
<td>22.74 (18–38)</td>
<td>3.6 years (1.5–7 years)</td>
<td>4.3 months (0–24 months)</td>
</tr>
<tr>
<td>B</td>
<td>8 (5/3)</td>
<td>21.38 (19–27)</td>
<td>3.9 years (2–10 years)</td>
<td>5 months (0–36 months)</td>
</tr>
<tr>
<td>C</td>
<td>19 (17/2)</td>
<td>21.21 (18–32)</td>
<td>4.7 years (1–10 years)</td>
<td>4.4 months (0–24 months)</td>
</tr>
<tr>
<td>D</td>
<td>8 (3/5)</td>
<td>20.75 (20–22)</td>
<td>4.4 years (2–7 years)</td>
<td>2.1 months (0–10 months)</td>
</tr>
<tr>
<td>E</td>
<td>10 (6/4)</td>
<td>21.7 (19–28)</td>
<td>4.2 years (3–7 years)</td>
<td>4.3 months (0–14 months)</td>
</tr>
<tr>
<td>F</td>
<td>13 (9/4)</td>
<td>21.31 (19–23)</td>
<td>3.3 years (2–7.5 years)</td>
<td>1.6 months (0–10 months)</td>
</tr>
<tr>
<td>G</td>
<td>23 (15/8)</td>
<td>23.48 (19–40)</td>
<td>3.1 years (0.75–9 years)</td>
<td>8.3 months (0–36 months)</td>
</tr>
<tr>
<td>H</td>
<td>10 (3/7)</td>
<td>25.6 (22–31)</td>
<td>3.9 years (1.75–11 years)</td>
<td>10.8 months (0–25 months)</td>
</tr>
<tr>
<td>I</td>
<td>16 (10/6)</td>
<td>20.5 (18–25)</td>
<td>2.0 years (1–4 years)</td>
<td>1.1 months (0–10 months)</td>
</tr>
<tr>
<td>Entire group</td>
<td>143 (89/54)</td>
<td>22.21 (18–40)</td>
<td>3.7 years (0.75–11 years)</td>
<td>4.7 months (0–36 months)</td>
</tr>
</tbody>
</table>

Of the 143 participants, 89 were male and 54 were female. The participants’ ages varied, but most were in their early twenties. The mean age was 22.21. The majority of the participants at each school were undergraduate students, except for those at School H in Japan, where all of the participants were graduate students or had graduated from university just before participating in the summer program. The participants had diverse academic backgrounds, and their majors included not only Japanese and Asian studies,
but also engineering, biology, economics, history, and so on. The mean length of studying Japanese was 3.7 years, ranging from 0.75 to 11 years. Regarding the length of time studying and living in Japan, a length of less than one month was calculated as zero. Half of the participants (71 participants) answered that they had never been to Japan or had stayed in Japan for less than one month. The mean length of stay in Japan was 4.7 months, ranging from 0 to 36 months.

Regarding participants’ first language (L1), Table 2 summarizes the information.

<table>
<thead>
<tr>
<th>School</th>
<th>L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>English: 32, Korean: 2, Chinese: 1, Russian: 1</td>
</tr>
<tr>
<td>B</td>
<td>English: 6, English &amp; Albanian: 1, English &amp; Spanish: 1</td>
</tr>
<tr>
<td>C</td>
<td>English: 15, Korean: 2, Arabic: 1, English &amp; Spanish: 1</td>
</tr>
<tr>
<td>D</td>
<td>English: 7, Korean: 1</td>
</tr>
<tr>
<td>E</td>
<td>English: 9, Korean: 1</td>
</tr>
<tr>
<td>F</td>
<td>English: 5, Korean: 4, Chinese: 1, Malay: 1, Russian: 1, English &amp; Korean: 1</td>
</tr>
<tr>
<td>G</td>
<td>English: 19, Chinese: 1, Serbian: 1, Vietnamese: 1, English &amp; Spanish: 1</td>
</tr>
<tr>
<td>H</td>
<td>English: 8, Chinese: 2</td>
</tr>
<tr>
<td>I</td>
<td>English: 6, Chinese: 5, Korean: 1, Spanish: 1, English &amp; Spanish: 2, English &amp; French: 1</td>
</tr>
</tbody>
</table>

107 of the participants (74.8%) were native speakers of English, whereas 28 participants (19.6%) reported that they were native speakers of other languages: 11
Korean speakers, 10 Chinese speakers, two Russian speakers, and one speaker each of Arabic, Malay, Serbian, Spanish, and Vietnamese. Eight participants (5.6%) described themselves as having two native languages including English. The other native languages of these participants were Spanish (5 participants), Albanian (1 participant), French (1 participant), and Korean (1 participant).

Although participants in some studies had a variety of L1 backgrounds as in this study (Elder & Iwashita, 2005; Ellis, 1987; Foster & Skehan, 1996; Mehnert, 1998; Skehan & Foster, 1997, 2005), most of the past planning studies involved participants with the same L1. However, none of the studies suggested a specific influence of the learners’ L1 on their L2 performance pertinent to task planning. One of the challenges to investigate the acquisition of less commonly taught languages, such as Japanese, is collecting enough data to obtain statistically reliable results. This is especially true when recruiting learners at higher levels. Considering the obstacles of recruiting a sufficient number of participants, the number of participants was prioritized over consistency in their L1, and therefore, the variation in their L1 was disregarded in this study.

The participants in this study were volunteers who received modest compensation upon completing their participation.

3.3 Materials

Four types of materials were involved in this study: a questionnaire, Simple Performance-Oriented Test (SPOT), materials for narrative tasks, and materials for the retrospective interviews. This section explains what these materials are and how they were developed or selected. Two pilot tests were conducted to determine the materials for this study: one for two picture sets for narrative tasks and the other for interview questions. These pilot tests are also illustrated in this section.
3.3.1 Questionnaire

The primary purpose of the questionnaire was to solicit information about Japanese learners’ past and current experiences with the Japanese language and determine their eligibility to participate in this study (Appendix A). It started with a section to collect the Japanese learners’ contact information and general information such as their first language, gender, and age, as well as their undergraduate major fields of study. The latter part of the questionnaire asked five questions. These questions were aimed at identifying Japanese learners who have had experiences that could affect the results and, therefore, would be eliminated from this study. The first question asked how long the Japanese learners had been studying Japanese. The second question was whether or not the learners had been to Japan and, if so, when, for how long, and for what purpose. The purpose of this question was to identify the learners who had lived in Japan for an extensive amount of time or those who had lived in Japan in their childhood. The third question asked whether they had a native speaker of Japanese in their immediate family and, if so, what their relationship to this person was and how often they spoke to the person in Japanese. The fourth question asked whether the learners spoke Japanese regularly outside of their Japanese class. If the answer was yes, they were further asked in what capacity and how often they spoke in Japanese. The third and fourth questions were used to exclude the Japanese learners who had close contact with a native speaker of Japanese (e.g., parent, spouse, friend, roommate) and who spoke Japanese with him/her outside of classroom frequently. Finally, the questionnaire asked whether the Japanese learners were regularly involved in speaking activities, such as speech and debate, either in the L1 or in the L2, to identify those who might have had extensive training in speaking.
3.3.2 SPOT

The Simple Performance-Oriented Test (SPOT) is a proficiency test for Japanese learners. This study adopted SPOT as a means of the necessary evaluation of participants’ Japanese proficiency levels for this study. The participants in this study were from nine different institutions, and although several considerations were taken to minimize the differences among the participants, it was undeniable that they had diverse backgrounds. For this reason, their Japanese abilities needed to be confirmed to ensure that they were at comparable proficiency levels. SPOT scores were also used to control the participants’ Japanese proficiency upon statistical analysis of the data.

SPOT was created by a group of researchers at the University of Tsukuba in Japan in the early 1990s. It was designed to evaluate real-time processing ability of Japanese grammar based on the hypothesis that language learners are able to comprehend linguistic forms that they already know, but not those they do not know (Ford, Kobayashi, & Yamamoto, 1995; Ford-Niwa & Kobayashi, 1999). SPOT was selected for this study because it has been empirically proved by past studies that its scores have a high correlation with various assessments that evaluate individual aspects of L2 learners’ Japanese abilities, including speaking (Hatasa & Tohsaku, 1997). Another advantage of using SPOT is that it is time-efficient, requiring approximately 10 to 15 minutes for the entire procedure.

SPOT consists of a series of unrelated sentences, which are recorded at normal speed on an audio tape (Ford, Kobayashi, & Yamamoto, 1995; Ford-Niwa & Kobayashi, 1999; Hatasa & Tohsaku, 1997). On the SPOT answer sheet, all of the sentences on the tape are written out, except for one hiragana, Japanese syllabary, in each sentence. The missing hiragana represents part of grammatical form. Test takers are supposed to fill in the blank in each sentence with one hiragana to complete the sentence while listening to the same sentence on the tape. Thus, the task of the test takers is to read each sentence on the answer sheet as they listen to the tape, identify the missing hiragana in the sentence,
and write it in a blank on the answer sheet. The interval between sentences is approximately two seconds, and, therefore, the processing of grammar has to be immediate for the test taker to write down the hiragana quickly enough to be ready for the next sentence.

Hatasa and Tohsaku (1997) summarized the characteristics of SPOT compared to a cloze test and a c-test as follows:

1. SPOT uses an audio tape, and the answer is given on the tape. Neither the cloze test nor c-test uses a tape.

2. SPOT items are independent from each other because each item appears once in a sentence and sentences in the test are all unrelated. On the other hand, in cloze and c-tests, test items are dependent on each other because more than one item is likely to be embedded in a sentence which is in the text.

3. SPOT is not affected by topic familiarity or text difficulty, unlike a cloze or c-test.

4. All deleted items in SPOT are grammar items. Items on a cloze test with a rational deletion procedure may or may not be grammar items, and items on a cloze test with a random deletion procedure or those in a c-test are not necessarily grammar items.

5. A letter which represents one mora is deleted in SPOT. One word or a part of a word is deleted in a cloze or c-test.

6. SPOT is a timed test in that learners have a limited time to fill in the blanks. A cloze or c-test is not necessarily a timed test.

(Hatasa & Tohsaku, 1997, p. 78)

An audio tape, answer sheets, and an instruction sheet for versions 2 and 3 of SPOT were obtained from the creator of SPOT in November, 2006. Version 2 was selected for this study because it was meant to assess intermediate and advanced levels of Japanese (Ford, Kobayashi, & Yamamoto, 1995). Version 2 of SPOT consists of 60 sentences and 5 short conversations. One hiragana is missing in each sentence and in
each set of conversations, for a total of 65 blanks. For each correct answer, 1 point was assigned, and therefore, 65 points was the maximum score.

3.3.3 Materials for the narrative tasks

Following many of the previous task planning studies (Crookes, 1989; Elder & Iwashita, 2005; Foster & Skehan, 1996; Gilabert, 2007; Kawauchi, 2005a, 2005b; Mochizuki & Ortega, 2008; Ortega, 1999; Park, 2006; Skehan & Foster, 1997; Tavokoli & Skehan, 2005; Yuan & Ellis, 2003), this study used oral narrative tasks with sets of pictures to examine L2 learners’ oral production. Narrative tasks are considered to be more cognitively demanding than some other tasks such as personal stories (Skehan & Foster, 1997) and, therefore, it is expected that benefits of planning opportunities on task performance would be more apparent (Kawauchi, 2005a).

There were some desirable attributes for narrative tasks used for this study to have. First of all, the tasks should be easy to understand to prevent learners from misinterpreting the story or being confused. At the same time, the story should not be too simple, so that participants would create stories involving a variety of vocabulary and grammar and would be able to exhibit their L2 abilities. Also, the story should be enjoyable so that participants would be motivated to tell it.

Two narrative tasks were needed for this study: a familiarization task (warm-up only; data were not included in the analysis) and a main task. A set of six pictures was used for each task to require participants to retell the story based on those pictures. To choose the most appropriate pictures for this study, five sets of pictures were selected for a pilot test. Three of the five sets were obtained from published sources for teaching foreign languages (The College Board, 1994; Ogiwara, Masuda, Saito, & Ito, 2005). Each set consisted of six pictures. The remaining two sets were created for this study. A teaching-assistant for the Japanese program at the University of Iowa and the researcher exchanged ideas for possible story lines that could be described in six pictures like the
existing three picture sets, and the teaching-assistant drew pictures based on the discussion. Each of the five picture sequences involved an accidental incident that was resolved at the end.

Two advanced-level Japanese learners who were not included in the operational study participated in the pilot test to decide on the picture sets for this study. They narrated a story for each set of pictures. After narrating all five stories, the participants provided feedback on the clarity of the pictures and difficulty of retelling the stories, and it was confirmed that none of the picture sequences was unclear, difficult to understand, or otherwise confusing.

All of the narratives created by the pilot study participants were transcribed, and different words in each narrative were counted to compare the variations in vocabulary. The variations in grammatical structures in the narratives were also measured. The mean number of different words used by the two pilot study participants was similar across the five narratives, and, therefore, the picture set that generated the most variation in grammatical structures was chosen for the main task, and the other one that generated the second most grammatical variation was selected for the familiarization task. Both picture sets can be seen in Appendix B.

For each of the familiarization and main tasks, an instruction sheet in English was prepared, so it was clear for the participants to know what they were supposed to do (Appendix C). The instruction introduced the name of the main character and instructed the participants under what task condition they were to recount what happened to the main character. Regardless of the assigned task condition, the participants were encouraged to describe the story in as much detail as possible as if they were talking to someone who had not seen the pictures, so that they would produce longer speeches for analysis.
3.3.4 Materials for the retrospective interviews

This study involved retrospective interviews, including stimulated recall, to elicit the information about the internal thought processes that the Japanese learners experienced when planning for a task. It is only possible through learners’ verbal reports about their own thoughts to gain access to their cognitive activities during planning and to find out how the opportunity to plan for the task affected their oral production. By analyzing the participants’ recalls on their thoughts while being engaged in planning, this study attempted to reveal whether there were any differences in their strategy use during planning depending on the type of planning they were allowed to have.

Prior to the interviews, a set of questions was prepared based on the interview questions in Ortega’s (1999) study (Appendix D). The purpose of pre-determined questions was to avoid directive questions and eliminate the influence of possible researcher’s biases on informants’ answers as much as possible. Asking the same questions was intended to make the type of information that could be obtained from the interviews consistent over the course of data collection.

The interview questions started with warm-up questions, followed by the questions regarding pre-task planning for the participants who were allowed to have planning time before task performance. The purpose of these questions about pre-task planning was to uncover participants’ cognitive activities to prepare for the task. The interview involved some questions about the planning during the task performance regardless of the planning conditions they had, to find out their on-going thought processes as they told the story. The group of participants who had a time limit for their task performance was asked whether the time pressure influenced their task performance in any ways. Questions during the stimulated recall were limited to what the participants were doing or thinking so that the participants would not be led to certain answers and so that the influence of researcher interference on the data would be minimized. Follow-up questions for participants’ answers were meant solely to clarify the participants’
explanations and gain detailed information about their internal activities. The interview ended with two general questions that provided the participants a final chance to add any information about their thoughts during planning and the task performance.

For the stimulated recall portion of the interview, the written instruction was created based on the sample instruction in Gass and Mackey (2000), so that it would be clear to the participants what they were expected to do during the recall. The instruction sheet explained that the objective of the stimulated recall was to learn what they were thinking specifically at the time of planning for the task as well as while telling the story, not at the time of the interview. In these instructions, the participants were requested to pause the video whenever they had something to say about their experiences or thoughts at the moment depicted in the video. The instruction for the stimulated recall also mentioned that the video might be paused by the researcher to ask them questions. To obtain more accurate data, it was important to prevent the participants from feeling pressed to say something and providing information about which they were uncertain. Hence, the instruction sheet included a statement that they could answer, “I don’t remember,” if they were not sure what they were thinking at the time of the video recording.

To determine the appropriateness of the interview questions and the structure of the interview, a second pilot test was conducted with three pre-intermediate students at School A. The participants met the researcher individually and performed the main task that was selected by the first pilot study described in section 3.3.3. They were assigned to different planning conditions to tell stories so that all of the interview questions could be tested. The planning conditions that were involved in this pilot study were as follows:
Condition 1: 10-minute pre-task planning time only

Condition 2: 10-minute pre-task planning time, a time limit of one minute and 30 seconds to perform the task

Condition 3: 10-minute pre-task planning time, a prompt every 30 seconds to encourage on-line planning during the task performance

All of the participants had 10 minutes to prepare for the task prior to the task performance. One of the participants had a time limit to perform the task, whereas another participant was told at 30-second intervals to think about what he/she was going to say during the task to encourage on-line planning. The participants were recorded on video during the planning time before the task and also during the task performance.

Immediately after completing the task, each participant was interviewed about his or her experience of both the planning and the narrative task following the list of questions that had been created. One of the participants did not understand one question in the list and, therefore, this question was modified after the pilot study. The rest of the questions elicited the information from the participants as intended. Following the interviews, the participants were asked about the clarity and coherence of the interview questions. Except for the question that one participant did not understand, it was confirmed that the questions were clear and easy to follow. The list of interview questions was finalized after the pilot test as shown in Appendix E.

3.4 Task conditions

This study involved pre-task planning and on-line planning, and four types of planning conditions were created by combining them: no planning, pre-task planning, on-line planning, and both pre-task and on-line planning. Participants under the no planning condition were given 30 seconds to look at the pictures for a task before telling a story. This amount of time was expected to be just enough for them to familiarize themselves
with the pictures and understand the story, but not so much that they were able to plan what they would say.

Under the pre-task planning condition, participants were allowed five minutes to prepare for the familiarization task and 10 minutes to prepare for the main task. They were provided with a sheet of paper to take notes during the planning phases, which was taken away before they started performing the tasks so that they would focus on telling a story instead of reading from their notes. The participants were instructed not to write down everything they were going to say either in Japanese or in English. Other than that, they received no instructions on how to plan for the task or what to write on the paper. It was up to the participants on how they would use their time before the task.

In regard to the on-line planning condition, Yuan and Ellis (2003) attempted to manipulate the amount of on-line planning by incorporating time pressure during task performance. They set a time limit for the no on-line planning groups to perform a task in an attempt to reduce to zero the amount of time available for on-line planning for these groups, whereas the participants in the on-line planning group were allowed to spend as much time as they wished to perform a task. However, it is not certain whether or not the participants planned on-line just because they had no time pressure for their task performance. In this study, three steps were taken to ensure that the participants under the on-line planning condition would truly engage in planning as they spoke. First, in the written instructions, they were told to think about what they were going to say while performing the task. Then, during the familiarization task, the participants were prompted to plan on-line by the researcher who said, “don’t forget that you can take time to think about what you are going to say” every 30 seconds, except when they paused and/or appeared to be thinking at the 30-second mark. If that was the case, another 30 seconds were given for the next prompt. Last, right before the participants began performing the main task, they were reminded to take as much time as they needed to think about the story while they were talking. The prompt was not provided during the main task to avoid
distracting the participants from their performance. Just as for those under the pre-task planning condition, the participants in the on-line planning group received no instruction on how to plan for the task.

To investigate the effects of time pressure on L2 oral production, a time limit was set to complete a task. The participants in the first pilot study, described earlier in this chapter, spent one minute and 40 seconds on average to tell their stories for the pictures that were selected for the familiarization and main tasks. Therefore, the time limit to perform these tasks was set at one minute and 30 seconds so that participants in this study who were assigned to the time pressure condition would actually feel that they were pressed for time during the task performance. Since the participants with on-line planning were encouraged to take time to think about what they would say while telling the story, it was contradictory to put time pressure on them. Also, the amount of on-line planning may be affected by time pressure as Yuan and Ellis (2003) claim. For these reasons, only those groups with no on-line planning opportunities were further divided into the groups with or without time pressure. As a result, six experimental groups were created: (a) no pre-task planning, no time pressure (control group, G1); (b) no pre-task planning, time pressure (G2); (c) pre-task planning, no time pressure (G3); (d) pre-task planning, time pressure (G4); (e) no pre-task planning, on-line planning (G5); and (f) pre-task and on-line planning (G6). The task condition for each group is shown in Table 3, followed by summaries of the conditions.
Table 3. Task conditions

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-task planning</th>
<th>Prompt for on-line planning</th>
<th>Time pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. No pre-task planning, no time pressure</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2. No pre-task planning, time pressure</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3. Pre-task planning, no time pressure</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4. Pre-task planning, time pressure</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>5. No pre-task planning, on-line planning</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>6. Pre-task and on-line planning</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Group 1: No pre-task planning, no time pressure (control group, G1)
Participants in the control group were allowed to spend 30 seconds looking at the pictures before task performance. They had no other preparation opportunity, and they had no time limit to complete the task.

Group 2: No pre-task planning, time pressure (G2)
Participants in this group also started performing the task 30 seconds after receiving the pictures. Unlike the participants in the G1, the participants in this group had a time limit of one minute and 30 seconds to complete the task.

Group 3: Pre-task planning, no time pressure (G3)
This group had a 10-minute preparation period prior to performing the task. The participants were provided with a sheet of paper to write notes, but they were instructed not to write out everything they intended to say either in Japanese or in English. They
were informed that the notes would be taken away as they began the task. They had no time limit to complete the task.

Group 4: Pre-task planning, time pressure (G4)

The participants in this group also had a 10-minute preparation period for the task with a sheet of paper for notes just like those in the PTP−T group. They were informed that the paper would be removed as they began the task. Unlike the G3, however, the participants in this group had a time limit of one minute and 30 seconds to complete the task.

Group 5: No pre-task planning, on-line planning (G5)

In this condition, participants were required to start narrating their story 30 seconds after receiving the pictures. They were told that they could take as much time as they would need to complete the task, and they were encouraged to think about the story while they speak. To remind the participants to plan while they tell their stories, the researcher told them “don’t forget that you can take time to think about what you are going to say” every 30 seconds during the familiarization task. The participants received no reminder of on-line planning during the main task.

Group 6: Pre-task and on-line planning (G6)

For this condition, participants were provided with both pre-task and on-line planning time. They had 10 minutes to prepare for the task with a sheet of paper to write notes. Like the other pre-tasking planning groups, they were informed that the notes would be removed as they began the task. The participants were allowed to take unlimited time to finish the task and were encouraged to think about the story while performing the task. Just as those in the G5, the participants in this group were prompted
every 30 seconds during the familiarization task to take time to think about the story, but did not receive the prompt during the main task.

3.5 Procedures

Participants were invited to meet with the researcher individually in a quiet room with recording equipment. First, the participants responded to the demographic questionnaire that is described in section 3.3.1. After the questionnaire, all of the participants took SPOT, except for 14 participants at School A, who had taken SPOT in their Japanese classes shortly before the data collection. While the participants were taking SPOT, the researcher reviewed their answers on the questionnaire. After completing SPOT, they were asked questions when clarifications for their questionnaire responses were needed.

Following the questionnaire and SPOT, the participants performed the familiarization task and the main task. Each participant was assigned to one of the six experimental groups to narrate stories for these tasks. Group assignment of 14 participants at School A, who took SPOT prior to the data collection, was based on their SPOT scores. These 14 participants were assigned across the six groups so that each group would have participants with comparable Japanese proficiency levels. Since the SPOT scores for the rest of the participants were unknown at the time of assigning them to groups, the assignment of task conditions had to be based on the best information available; that is, the Japanese course level in which they were enrolled in at their schools. As much as possible, the students at each course level in each school were distributed evenly across the six task conditions.

Prior to the familiarization task, the participants received the written instructions for the task and read the explanation about the setting of the narrative story and the planning condition to perform the task. After the participants confirmed that they understood the instructions, they received a set of pictures for the task and were asked to
retell the story based on those pictures. Upon completing the familiarization task, the participants were asked if they had any questions about the planning condition and the task, and, if they did, the researcher answered their questions. Then, the participants were given the written instructions for the main task (Appendix C) and performed the task under the same planning condition as with the familiarization task.

The participants in the G2 and G4 had a time limit of one minute and 30 seconds to perform the tasks. For both the familiarization task and the main task, they were reminded that they had a time limit to complete the task as they started telling the story, whereas the participants in the other groups were told right before the performance that they had no time limit for the task. If the participants in the G2 and G4 passed the time limit while performing the familiarization task, they were told how much time they had taken after finishing the story so that they would be aware of the time limit and would try to remain within it when performing the main task.

The participants in G3, G4, and G6 had an opportunity to plan for the tasks prior to the task performance. They were allowed five minutes to prepare for the familiarization task and 10 minutes for the main task. For each task, the participants in these groups were provided with a sheet of paper to take notes during the pre-task planning in addition to the set of pictures for the narrative task. When the planning time was over, the notes were taken away as they began performing the task. There were a few participants who told the researcher that they were ready to start telling the story before the assigned pre-task planning time had passed. In these cases, the researcher told them how much time was still remaining and instructed them to spend the rest of the time planning.

Regarding the on-line planning condition, the participants in G5 and G6 received verbal prompts that encouraged thinking as they told the story while performing the familiarization task (section 3.4). For the main task, they were reminded to engage in on-line planning again at the beginning of the main task, but not during the task.
The experiment took approximately 20 minutes for G2 (no pre-task planning, with time pressure) and 40 minutes for G6 (pre-task planning and on-line planning). The other experimental groups took anywhere between 20 to 40 minutes.

All of the participants were video-recorded during the task performance, and their narratives were later transcribed for analysis. Immediately after the main task, 24 participants in Schools A and B, who took part in the qualitative component of this study, received retrospective interviews in English to document their perceptions of how they had planned for the task and what they thought while telling the story. The video-recording for these participants started from the time they received the pictures for the main task regardless of the opportunity for pre-task planning and was continued through their task performance to keep the record of their thinking processes both before and during the task.

The retrospective interviews were given based on the questions in the interview sheet that was explained in section 3.3.4. As was described in section 3.3.4, the interview started with the questions that asked the participants about their experiences of planning and task performance (Appendix D). The second part of the interview was a stimulated recall of their thoughts before and during the task while watching participants’ own videos. Prior to the recall, the participants received the written instructions for the stimulated recall (Appendix E) and were told to report any moments when they remembered what they were thinking while planning for the task or while performing the task.

The procedures for the stimulated recalls were determined through the second pilot study that was described in section 3.3.4. For this pilot study, the video was stopped every 30 seconds during the pre-task planning phase. During the task performance the video was stopped every time the participants started talking about the next picture as well as when they paused. Each time the video was stopped, the participants were asked to report what they remembered thinking at the segment in the video. Following the
interviews, the participants in the pilot study were asked about their experiences of the interviews. Regarding the stimulated recall, two of the participants expressed that it was uncomfortable to watch themselves, particularly during the 10-minute planning time, because they stayed still most of the time while planning, and they had little to say to explain what they were thinking during the period. To avoid making participants feel awkward, it was decided to fast forward the video during the pre-task planning until it showed facial expressions or body movements that might reflect conscious changes in thinking. In other words, the decision was made to stop the video during the pre-task planning phase to ask participants’ thoughts when (a) they made distinctive actions, such as writing down notes or whispering something and (b) their eyes moved from one picture to another. The rest of the video for pre-task planning phase was fast forwarded.

The procedures for the stimulated recall were unchanged from the pilot study otherwise; specifically, the video during the task performance was shown at normal speed and was stopped at the segment where the participants paused or started talking about the next picture. The video was also stopped when the researcher had questions. In addition, the participants were encouraged to stop the video whenever they remembered something about their thinking processes at that time in the video both during planning and during task performance.

The length of the interviews ranged between 10 and 15 minutes for the participants without a pre-task planning opportunity (G1, G2, and G5) and between 20 and 30 minutes with a pre-task planning opportunity (G3, G4, and G6).

Figure 1 describes the procedures for this study.
Figure 1. Procedures

Demographic questionnaire

⇓

SPOT

⇓

Familiarization task

⇓

Main task

⇓

Retrospective interviews
(24 participants in Schools A & B only)
3.6 Measures

Two sets of measures were involved in this study. The first set of measures was taken to answer research questions 1 and 2, which investigate the effects of time pressure as well as planning on participants’ narrative stories in terms of fluency, complexity, and accuracy. The data source for the quantitative analyses was the recorded narrative stories for the main task. The second set of measures was taken to answer research question 3, which examined the participants’ strategy use while planning and performing the task. The data source for these analyses was audio recordings of the retrospective interviews. This section explains procedures of data transcription, followed by the description of each set of measures. The section also illustrates coding processes and the results of inter-rater reliability for the coding.

3.6.1. Measures for L2 oral production

3.6.1.1 Data source: Transcriptions of oral production

The data source for the analysis of L2 oral production was transcriptions of the participant’s narrative stories in Japanese. For the transcriptions, regular Japanese writing systems, including punctuation conventions, were used. In addition, an “X” was used to indicate an incomprehensible mora, the Japanese phonological unit. The explanation of mora is found in section 3.6.1.2.1. When participants used non-Japanese words, such as English, to tell stories, these parts were written in the Roman alphabet. Any utterances during the task performance that were not considered as part of the narrative stories (self-talk and questions to the researcher for instance) were enclosed in parentheses. Fillers in the participants’ oral production were excluded in the transcriptions to separate them from the intended speeches. Table 4 summarizes the transcription conventions used in this study.
Table 4. Sample transcription conventions

<table>
<thead>
<tr>
<th>Punctuation:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td>Japanese sentence end marker</td>
</tr>
<tr>
<td>,</td>
<td>Japanese comma</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Others:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>incomprehensible mora</td>
</tr>
<tr>
<td>( )</td>
<td>not part of narrative stories</td>
</tr>
<tr>
<td>alphabet</td>
<td>non-Japanese words</td>
</tr>
</tbody>
</table>

Since the narrative stories were quantified on several measures for statistical analyses, inaccurate transcriptions would directly affect the results of this study. To create the most accurate transcriptions, the following steps were taken to finalize the transcriptions of the task performance. First, a graduate student who is a native speaker of Japanese was recruited, and she and the researcher transcribed all of the participants’ narratives for the main task individually following the transcription scheme. Second, the two sets of transcriptions were compared, and when there were any discrepancies between them, these parts were included in the transcriptions for further verifications by listing two alternatives next to each other in parentheses divided by a slash mark as follows:
(1) そのコンピューターは「使えませんでした／使いませんでした」。

*sono conpuutaa wa [tsukae masen deshita/tsukai masen deshita].*

(He) [could not use/did not use] the computer.

In the example above, one person transcribed the phrase as “*tsukae masen deshita* (could not use)” and the other as “*tsukai masen deshita* (did not use).” To make these inconsistent parts distinctive, the font color for these parts was changed to blue, whereas the rest of the transcriptions was in black. The agreement rate between the two transcriptions in terms of mora was 96.49%.

Third, to resolve these discrepancies, another graduate student, who is also a native speaker of Japanese, was recruited to check the transcriptions. She watched the videos for the participants’ task performance and verified what participants had said by circling one of the two alternatives for each discrepancy. When either alternative did not match with what she thought she heard, she added the alternative in the transcriptions as the third option. Finally, after all of the inconsistent parts were confirmed by the second graduate student, the data were finalized by erasing the options in the transcriptions that were not selected. For each part for which a third option was added, the researcher watched the video again and determined which of the three options should be included in the transcriptions.

3.6.1.2 Measures

The following 12 measures were subjected to a factor analysis to determine most appropriate dependent variables for this study (Table 5).
Table 5. Measures for L2 oral production

<table>
<thead>
<tr>
<th>Fluency measures</th>
<th>Number of moras per minute</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of meaningful moras per minute</td>
</tr>
<tr>
<td></td>
<td>Number of pauses per minute</td>
</tr>
<tr>
<td>Complexity measures</td>
<td>Number of clauses per T-unit (syntactic)</td>
</tr>
<tr>
<td></td>
<td>Number of words per T-unit (syntactic)</td>
</tr>
<tr>
<td></td>
<td>Type-token ratio (lexical)</td>
</tr>
<tr>
<td></td>
<td>Modified type-token ratio: type/ √2*token (lexical)</td>
</tr>
<tr>
<td>Accuracy measures:</td>
<td>Percentage of error-free clauses (global)</td>
</tr>
<tr>
<td></td>
<td>Number of errors per 100 words (global)</td>
</tr>
<tr>
<td></td>
<td>Number of non-repeated errors per 100 words (global)</td>
</tr>
<tr>
<td></td>
<td>Percentage of correct use of particles (specific)</td>
</tr>
<tr>
<td></td>
<td>Percentage of correct verb forms (specific)</td>
</tr>
</tbody>
</table>

A description of each measure and of coding methods is presented in the following sections.

3.6.1.2.1 Fluency measures

To measure the fluency in the participants’ oral production, the numbers of moras, meaningful moras, and pauses within each narrative were counted. Many of the previous studies have used the number of syllables to measure fluency (Gilabert, 2007; Mehnert, 1998; Ortega, 1999; Sangarun, 2005; Tavokoli & Skehan, 2005; Wendel, 1997; Yuan & Ellis, 2003). However, this study used moras instead of syllables because the mora is considered to be the basic timing unit to indicate phonological length in Japanese.
Tsujimura (2007) explains that a mora can be formed in the following three ways:

1. (C)V
2. the first part of a long consonant (or the first part of a geminate)
3. syllable final, or “moraic,” nasal /n/  

(Tsujimura, 2007, p. 59)

As presented above, one mora may be represented by one vowel with or without a consonant preceding the vowel, such as /a/ and /ka/. With respect to consonants, there are two occasions when a consonant forms one mora by itself. One is when a long consonant occurs. A long consonant, which is also called a double consonant because two of the same consonants are written consecutively to represent this phenomenon, is a short pause that occurs between two sounds (Amanuma, Otsubo, & Mizutani, 1978). For a long consonant, the first of the two consonants is counted as one mora. For example, in /kitte/ (postal stamp) the first /t/ is considered to be one mora. Another occasion is when a nasal /n/ occurs without a following vowel. Therefore, /n/ in /hon/ (book) forms one mora as well. Japanese is considered a mora language because a syllable language requires a vowel to form a basic prosodic unit, and a consonant cannot form an independent unit by itself, as some consonants can in Japanese as defined in (b) and (c) (Kubozono, 1999; Tsujimura, 2007).

For the calculation of the number of moras in the speech data, the word count function in Microsoft Word was used to compute the number of kana (hiragana and katakana) in the transcriptions, since one mora generally corresponds to one kana, and therefore, the number of kana reflects the number of moras. To run the word count function correctly, the transcriptions were modified in the following order.
1) Write out the readings of Chinese characters in kana
2) Eliminate one kana from a Japanese paralyzed sound
3) Remove Japanese commas and periods

First, the transcriptions for this study were written in a combination of kana (hiragana and katakana) and Chinese characters, which is the conventional way of writing Japanese texts. To modify the transcriptions to suit the word count function, they were rewritten in kana only by writing out the readings of the Chinese characters in hiragana. Second, one kana each from a Japanese palatalized sound (e.g., /ky, kyo/) was removed from the transcriptions because they are considered as one mora, while they are written in a combination of two kana letters. Finally, since the word count function in the computer includes Japanese commas and periods in counting, they were removed from the transcriptions, so that only kana would remain in the transcriptions.

At times, participants in this study used English words to narrate their stories when they were not sure how to say something in Japanese. On these occasions, the number of syllables was counted and included in the data. The total number of syllables for these English words was 39, which was 0.09% of the data and considered small enough not to affect the results of the study.

After counting the number of moras, and syllables when English words were included, for each narrative story in the modified transcriptions, the speech rate was calculated by dividing the number of moras in the narrative by the amount of time in seconds to complete the story, which was multiplied by 60.

The second measure was the number of meaningful moras per minute. Meaningful moras exclude the parts of narrative stories that were repeated, reformulated, or replaced (Yuan & Ellis, 2003). Calculating the number of meaningful moras was intended to evaluate the speech rate for the narrative story itself, separate from that for the entire speech that might include a varying amount of redundancy. To compute the
number of meaningful moras per minute, the transcriptions with *kana* only were used. In the transcriptions, the parts that were considered repeated, reformulated, or replaced by the participants were identified and eliminated from the transcriptions. The number of remaining moras was counted for each narrative, and then divided by the number of seconds used to complete the task and was multiplied by 60.

The third measure for fluency was the number of pauses per minute. Tajima (2003) defined a pause as silence or fillers for the length of two seconds or longer. In her pilot study on task planning, Tajima (2003) at first set the time for a pause at one second or longer. The results showed no significant differences in the frequency of pauses between planned and non-planned oral productions. After examining the data, Tajima (2003) reported that one second might have been inefficient as the cut-off point to determine pauses because at many occasions the participants in her pilot study paused for one to two seconds at the end of phrases or units in their speeches, but did not seem to have hesitation or difficulties to speak. Rather, these pauses appeared to be the ones that the participants would do even when they were speaking in their L1. Based on these observations, Tajima (2003) set two seconds as a cut-off point for pauses, and this study employed the same length of time for the identification of pauses.

To code pauses longer than two seconds, the video recordings of the participants’ task performance were transformed into sound files by first being imported by iMovie and then being saved by QuickTime on a Macintosh computer. These sound files were edited to highlight pauses, using Audacity, an audio editing software. After counting pauses, the number of pauses was divided by the length of time in seconds for each narrative story to compute the number of pauses per minute and was multiplied by 60.

3.6.1.2.2 Complexity measures

The present study employed two measures each for syntactical complexity and lexical variety. Syntactic complexity was measured using the number of clauses per T-
unit and the number of words per T-unit in the narratives. A T-unit, or minimal terminal unit, is defined as “one main clause with all subordinate clauses attached to it” (Hunt, 1965, p. 20). It was originally created to measure L1 development in writing, and there are some criticisms for using the T-unit to analyze spoken data (Foster, Tonkyn, & Wigglesworth, 2000). Despite the criticisms, the T-unit has been widely used in L2 studies to examine oral production, including many on planning (Crookes, 1989; Ellis & Yuan, 2005; Gilabert, 2007; Kawauchi, 2005a, 2005b; Mehnert, 1998; Mochizuki & Ortega, 2008; Sangarun, 2005; Yuan & Ellis, 2003), because it is considered to be an ideal measure in terms of objectivity, accessibility, and reliability (Harrington, 1986; Sotillo, 2000). Harrington (1986) conducted a study to investigate whether the T-unit could be adopted to analyze L2 speeches in Japanese. Following the definition of T-unit by Hunt (1965), Harrington (1986) defined a Japanese T-unit as “a nuclear sentence with its embedded or related adjuncts” (p. 53) and explained that a nuclear sentence can stand independently or can be connected with other words by particles to form one T-unit. He found that the length measures, such as average number of words per T-unit and average number of words per error-free T-unit, successfully distinguished participants’ proficiency levels.

Based on Harrington’s definition of the Japanese T-unit, Iwashita (2006) also investigated the reliability of the T-unit for Japanese oral production by comparing several measures that involve the number of words, clauses, and T-units. Among the measures that Iwashita (2006) examined, the number of words per T-unit and the number of clauses per T-unit were found to be most reliable to assess proficiency levels in L2 Japanese. Therefore, these two measures were employed in this study.

For the assessment of complexity, clauses and T-units in the data were coded. The ratio of clauses to T-units in the participants’ narratives was calculated by dividing the number of clauses by the number of T-units in each narrative. For the second measure of syntactic complexity, the number of words per T-unit, first, the number of words for each
narrative was computed by ChaSen, a Japanese morphological parser that was created by researchers at Nara Institute of Science and Technology in Japan. After counting the number of words, it was divided by the number of T-units in the narrative.

For lexical complexity, type-token ratio (TTR) and modified TTR (type/√2*token) were used in this study. TTR has been widely used to measure lexical complexity, though it has been criticized for being sensitive to length of samples (Ortega, 1999; Tajima, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998). The modified TTR was created as an attempt to solve such a problem of TTR by taking sample sizes into consideration upon calculation (Tajima, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998). As explained earlier in this section, the number of words was calculated by using ChaSen. The number of word types was also calculated by ChaSen. TTR was calculated by dividing the number of word types by the number of words produced by each participant. The modified TTR was computed by having “the number of word types divided by the square root of two times the total number of words” (Wolfe-Quintero, Inagaki, & Kim, 1998).

3.6.1.2.3 Accuracy measures

To evaluate accuracy of L2 production, Ortega (1995) suggested employing both global and specific measures so that global measures would capture more general changes in accuracy (Skehan & Foster, 1997) whereas specific measures would detect subtle accuracy effects that global measures might overlook. This study used three global measures and two specific measures for evaluation of accuracy in narrative stories.

For global accuracy, the percentage of error-free clauses, number of errors per 100 words, and number of non-repeated errors per 100 words were employed. Of these global measures, the percentage of error-free clauses has been most commonly used in previous planning studies (Ellis & Yuan, 2005; Elder & Iwashita, 2005; Foster & Skehan, 1996; Mehnert, 1998; Sangarun, 2005; Skehan & Foster, 1997, 2005; Tajima, 2003;
Tavakoli & Skehan, 2005; Yuan & Ellis, 2003). Skehan and Foster (2005) pointed out a possible problem for percentage of error-free clauses—that it might be artificially increased if oral production consists of many short phrases. To avoid such an issue, the number of errors per 100 words was used, following Mehnert (1998) and Sangarun (2005). In addition, number of non-repeated errors per 100 words was used so that the results would not be affected by the re-occurring errors.

Errors could be syntactic, morphological, or lexical, including inappropriate use and erroneous omission of words and grammatical elements and incorrect word order. When participants repeated, reformulated, or replaced any part of their narratives, only the last one they said was taken into account to determine whether it was correct or not. Non-Japanese words were also counted as errors. At times, some participants used English words with Japanese-like pronunciation, and they were considered as errors as well.

For accuracy, all of the errors in the transcripts were coded. Error-free clauses were those that contained no error. When any part of a clause was inaudible, the clause was excluded from the calculation for the percentage of error-free clauses, since it was impossible to determine whether or not the clause included any errors. The percentage of error-free clauses for each narrative story was calculated by dividing the number of error-free clauses by the total number of clauses in the narrative.

\[
\frac{\text{Number of error-free clauses}}{\text{Total number of clauses in the narrative}} \times 100
\]

To calculate the number of errors per 100 words, the number of errors in each narrative was divided by the number that was obtained by dividing the number of words in the narrative by 100. For the number of non-repeated errors per 100 words, the same errors that occur for the second time or more were excluded from the total number of
errors. Then, the number of errors was divided by the same number that was used as a denominator to calculate the number of errors per 100 words.

For specific accuracy, past planning studies have examined various linguistic domains, such as correct verb forms (Ellis & Yuan, 2005; Wendel, 1997; Yuan & Ellis, 2003), past-tense markers (Ellis, 1987; Kawauchi, 2005a, 2005b), plural -s (Crookes, 1989; Wigglesworth, 1997), and articles (Crookes, 1989; Ortega, 1995a, 1999; Wigglesworth, 1997). Among these grammatical items, this study employed correct verb forms and examined the percentage of their accurate use.

To compute the percentage of error-free verb forms, the researcher first identify incorrect verbs from the errors that had been already coded to calculate the percentage of error-free clauses, and then, counted the number of these errors. The number of verbs that were not coded as incorrect was also counted as error-free verbs. The number of error-free verb forms was divided by the total number of verbs in each narrative, which was computed by adding the number of error-free verbs and incorrect verbs.

\[
\frac{\text{Number of correct verb forms}}{\text{Number of correct verb forms} + \text{number of incorrect verb forms}} \times 100
\]

In addition to verb forms, this study employed particles for a specific measure because Tajima (2003), the only past planning study that investigated the effects of planning on L2 oral production in Japanese, examined the accuracy of particles. Japanese particles are words that indicate the relationships among words in a sentence as well as relationships between sentences (Makino & Tsutsui, 1995). Similar to English prepositions, particles are not conjugated and cannot stand by themselves. They are accompanied by other words, such as nouns, verbs, and adjectives but, unlike English prepositions, Japanese particles are attached after these words (Iori, Takanashi, Nakanishi, & Yamada, 2000). Table 6 shows the particles that were examined in this study.
Table 6. Particles in this study

<table>
<thead>
<tr>
<th>Particles</th>
<th>Meaning/functions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>wa</strong></td>
<td>1. Marking a topic or a contrastive element</td>
<td>Watashi wa gakusei desu. (I am a student.)</td>
</tr>
<tr>
<td></td>
<td>2. Indicating the subject</td>
<td>Ame ga futte imasu. (It’s raining.)</td>
</tr>
<tr>
<td></td>
<td>2. Marking the direct object for transitive adjectives or stative transitive verbs</td>
<td>Boku wa supootsukaa ga hosii. (I want a sports car.)</td>
</tr>
<tr>
<td><strong>ga</strong></td>
<td>1. Indicating the subject</td>
<td>Ame ga futte imasu. (It’s raining.)</td>
</tr>
<tr>
<td></td>
<td>2. Marking the direct object for transitive adjectives or stative transitive verbs</td>
<td>Boku wa supootsukaa ga hosii. (I want a sports car.)</td>
</tr>
<tr>
<td><strong>o</strong></td>
<td>1. Marking a direct object</td>
<td>Nani o nomimasu ka. (What will you drink?)</td>
</tr>
<tr>
<td></td>
<td>2. Indicating a space in/on/through/along which somebody or something moves</td>
<td>Kooen o tootte kaerimashoo. (Let’s go home through the park.)</td>
</tr>
<tr>
<td></td>
<td>3. Marking the location from which some movement begins</td>
<td>Watashi wa asa shichijihan ni uchi o demasu. (I leave home at 7:30 in the morning.)</td>
</tr>
<tr>
<td><strong>ni</strong></td>
<td>1. Indicating a point of time at which something takes place</td>
<td>Watashi wa maiasa mokujihan ni okimasu. (I get up at six thirty every morning.)</td>
</tr>
<tr>
<td></td>
<td>2. Indicating an indirect object</td>
<td>Watashi wa haha ni tegami wo yoku kakimasu. (I often write letters to my mother.)</td>
</tr>
<tr>
<td></td>
<td>3. Indicating an agent of a source in passive, causative, and receiving constructions</td>
<td>Watashi wa chichi ni kuruma wo katte moraimashita. (My father bought me a car.)</td>
</tr>
<tr>
<td></td>
<td>4. Indicating purpose when something moves from one place to another</td>
<td>Watashi wa depaato e okurimono wo kai ni ikimashita. (I went to a department store to buy a gift.)</td>
</tr>
<tr>
<td></td>
<td>5. Indicating the location where somebody or something exists</td>
<td>Sono hon wa kono gakkoo no toshokan ni arimasu. (That book is in this school’s library.)</td>
</tr>
<tr>
<td></td>
<td>6. Indicating a place toward which somebody or something moves</td>
<td>Jumu wa rainen America ni kaeru. (Jim is going back to America next year.)</td>
</tr>
</tbody>
</table>
### Table 6. Continued

|   | 1. With a preceding noun phrase, forming a phrase to modify a following noun phrase | Kore wa sensee no hon desu.  
(This is my teacher’s book.) |
|---|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| no | 2. Indicating the subject or object of a verb or adjective in a noun phrase       | Tenki no ii hi ni wa yoku sanpo shimasu.  
(I often take a walk on days when the weather is nice.) |
|   | 1. Indicating the place where an action occurs                                  | Hirugohan wa doko de tabemashoo ka.  
(Where shall we eat lunch?) |
| de | 2. Pinpointing a time period or physical space                                  | Sekai de ichiban takai yama wa Eberesuto desu.  
(The highest mountain in the world is Mount Everest.) |
|   | 3. Indicating the use of something for doing something                          | Watashitachii wa nihongo de hanashita.  
(We talked in Japanese.) |
|   | 4. Indicating boundaries of time, value, quantity                               | Kono tokei wa ichiman-en de kaimashita.  
(I bought this watch for ten thousand yen.) |
|   | 1. Indicating the direction toward which some directional movement or action proceeds | Kinoo ginkoo e ikimashita.  
(I went to the bank yesterday.) |
| e  | 1. Indicating a starting point or a source                                     | Paati wa hachiji kara hajimarimasu.  
(The party starts at eight o’clock.) |
| kara| 1. Indicating a spatial, temporal or quantitative limit                         | Koko kara eki made nanpun gurai desu ka.  
(How many minutes is it from here to the station?) |


Among the Japanese particles, *wa*, *ga*, *o*, *ni*, *no*, *de*, *e*, *kara*, and *made* were selected for analysis because they are usually introduced in elementary-level Japanese textbooks (Iori et al., 2000, Hatasa, Hatasa, & Makino, 2000; Makino, Hatasa, & Hatasa, 1998), and it was expected that all the participants had formally learned these particles by the time of the data collection.

For the errors of the Japanese particles (*wa*, *ga*, *o*, *ni*, *no*, *de*, *e*, *kara*, and *made*), following Tajima (2003), this study defined non-target-like use of particles as “erroneous use, deletion, and overuse” (p. 49). The data were coded for particles that were correctly used and those that were incorrectly used or omitted. To calculate the accuracy rate for particles, the number of correctly used particles was divided by the number of particles produced as well as the number of particles erroneously omitted.

\[
\frac{\text{Number of correctly used particles}}{\text{Number of particles produced} + \text{Number of particles erroneously omitted}} \times 100
\]

3.6.1.3 Interrater reliability

To establish interrater reliability for coding for the task performance data, the same graduate student who transcribed the participants’ narrative stories was trained on the coding system. The data were coded for the following six categories: extra moras, clauses, T-units, correct use of particles, erroneous use of particles, and other errors. At the beginning of the training session, the researcher gave instructions on the coding system to the rater, and the rater and the researcher completed the coding of three randomly selected narratives together. The rater had opportunities to ask questions for clarification and she and the researcher discussed the rationales for the coding decisions. Following the training, the rater and the researcher coded three transcriptions individually, and the two sets of coding were compared. The interrater reliability for each category was
calculated, and all of them exceeded the 95% level. Then, the rater and the researcher coded 29 data samples (20.28% of the data). Interrater reliability was calculated for two sets of coding and was established at a 90% or above agreement rate for each category, ranging 90% to 98.21% (Table 7).

Table 7. Interrater reliability for 29 data samples

<table>
<thead>
<tr>
<th>Categories</th>
<th>Extra moras</th>
<th>Clauses</th>
<th>T-units</th>
<th>Correct particles</th>
<th>Erroneous particles</th>
<th>Other errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrater reliability</td>
<td>94.27%</td>
<td>96.30%</td>
<td>96.44%</td>
<td>98.21%</td>
<td>90%</td>
<td>93.04%</td>
</tr>
</tbody>
</table>

3.6.2. Evaluation of learner strategies

3.6.2.1 Strategy classifications

All of the retrospective interviews were given in English and were audio-recorded. The interview data were transcribed by a native speaker of English. To code the data for learner strategies, this study employed classifications in Ortega’s (2005) study, which was developed based on O’Malley and Chamot (1990) and Oxford (1990). Ortega (2005) classified the strategies that the participants used in her study for three categories of learner strategies: metacognitive, cognitive, and social/affective strategies. Oxford (1990) defines metacognitive strategies as “actions which go beyond purely cognitive devices, and which provide a way for learners to coordinate their own learning process” (p. 136). Metacognitive strategies include planning, monitoring, and evaluation (O’Malley & Chamot, 1990). Cognitive strategies involve a wide range of strategies that directly deal with the target language such as rehearsal, approximation, and summarizing (O’Malley &
Chamot, 1990). On the other hand, social/affective strategies manage learners’ emotional aspects by inducing relaxation and providing encouragement and also include concerns for other people (Oxford, 1990).

Each strategy category was further divided into several sub-categories. Some strategies that were not adequate for analyzing the data for this study due to the differences in study designs between Ortega (2005) and this study were eliminated. Tables 8, 9, and 10 show the strategy categories for this study and the definition and example for each strategy.
### Table 8. Definitions and examples of metacognitive strategies

<table>
<thead>
<tr>
<th>Strategy categories</th>
<th>Definitions &amp; Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Planning</td>
<td></td>
</tr>
</tbody>
</table>
| 1.1. Advance planning | a) Previewing the main ideas and concepts of the material to be learned, often by skimming the text for the organizing principle (Advance organization, p. 229).  
  e.g.) G3–3: Like at first, I kind of ran through it very quickly, and now probably just looking at it more carefully. |
| 1.2. Organizational planning | a) Generating a plan … to be used in a language production task (p. 231).  
  e.g.) G6–1: So the first thing occurred to me was that I need to try to figure out the setting and who is talking. |
| 2. Problem identification | a) Explicitly identifying the central point needing resolution in a task, or identifying an aspect of the task that hinders its successful completion (p. 231).  
  e.g.) G3–1: There I knew I was going to have trouble with this one. [Interviewer: The fourth picture?] Yeah, because the cord, I didn’t know how to say ‘to plug in,’ so the whole time I was just thinking. |
| 3. Self-monitoring |                        |
| 3.1. Production monitoring | a) Checking, verifying, or correcting one’s language production (p. 137).  
  e.g.) G5–3: I think I was just still concentrating on worrying about grammar and things like that and trying to form correct sentences. |
| 3.2. Auditory monitoring | a) Using one’s “ear” for the language (how something sounds) to make decisions (p. 137).  
  e.g.) G3–2: I caught myself in saying sorede for the third time, so I was like, “Oh, I should use something else.” |
| 3.3. Visual monitoring | a) Using one’s “eye” for the language (how something looks) to make decisions (p. 137).  
  e.g.) G4–3: Here’s where I’m starting to write down what I think might be the word for “to turn on” as in electricity, and I wrote down a couple of words, none of them which I thought were correct, so I just crossed them out, and then figured I would start thinking of a different word. |
Table 8. Continued

<table>
<thead>
<tr>
<th>3.4. Cross-language monitoring*</th>
<th>c) a conscious comparison of the expression of an idea in the L1 and L2 in order to judge the ‘transferability’ of an L1 utterance (Ortega, 2005, p. 103).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e.g.) [I: Can you give me an example of trying to remember your grammar?] Oh, when to use subjunctive, you know, when they tell him to leave you know to say <em>vayan</em> {subjunctive imperative} instead of something else, or to put le {him, preverbal clitic pronoun} in front of <em>dice</em> {tells} and all those things that I — you don’t do in English, that you have to start to think, they don’t come automatically (Ortega, 2005, p. 96).</td>
</tr>
<tr>
<td>3.5. Style and register monitoring**</td>
<td>d) Checking, verifying, or correcting style and register of one’s speech.</td>
</tr>
<tr>
<td></td>
<td>e.g.) G3−1: There I was going to say “<em>otoko no hito no sensei,</em>” but then I was thinking that doesn’t sound right, or maybe it’s not polite enough. So I thought it’d be more polite in my explanation to say “<em>dansei.</em>”</td>
</tr>
<tr>
<td>3.6. Double-check monitoring</td>
<td>a) Tracking, across the task, previously undertaken acts or possibilities considered (p. 137).</td>
</tr>
<tr>
<td></td>
<td>e.g.) I was mouthing different possibilities for phrases that I didn’t know how to say, and I wrote out what I was saying, and trying to remember also the story in English. Then I looked at everything to see if it fits correctly. And then I guess I was telling it to myself in Spanish to see how it sounds, I reread it a couple of times and rehearsed it (Ortega, 2005, p. 96).</td>
</tr>
<tr>
<td>3.7. Strategy monitoring**</td>
<td>a) Tracking use of how well a strategy is working (p. 137).</td>
</tr>
<tr>
<td></td>
<td>e.g.) G5−2: I …tried to make it too complicated the first time I looked at it. I tried to be like, “I should say how many computers there are,” or make up names for people, or “What was the expression on this person’s face,” but …. Then I’d be like, “Well, that wasn’t necessarily crucial to the story,” I thought, “Well, the part of the story that’s important is that his computer wasn’t plugged in, I should focus on that.”</td>
</tr>
</tbody>
</table>
Table 8. Continued

<table>
<thead>
<tr>
<th>4. Self-evaluation</th>
<th>a) Judging one’s overall execution of the task. (p. 137).</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1. Performance evaluation</td>
<td>e.g.) G3−2: [I: Do you remember what you were thinking while you were talking about the sixth picture?] I thought that I got it about right.</td>
</tr>
<tr>
<td>4.2. Repertoire evaluation</td>
<td>a) Judging how much one knows of the L2 at the word, phrase, sentence, or concept level (p. 138).</td>
</tr>
<tr>
<td></td>
<td>e.g.) I used <em>tsukeru</em>, but I was wondering if it was right. I knew that you use <em>tsukeru</em> for light, but I wasn’t sure if I could use it for computer, too.</td>
</tr>
<tr>
<td>4.3. Ability evaluation</td>
<td>a) Judging one’s ability to perform the task (p. 137).</td>
</tr>
<tr>
<td></td>
<td>e.g.) I don’t know enough Japanese to describe this picture.</td>
</tr>
<tr>
<td>4.4. Strategy evaluation</td>
<td>a) Judging one’s strategy use when the task is completed (p. 137).</td>
</tr>
<tr>
<td></td>
<td>e.g.) G3−1: I was thinking that I should have written a sentence instead.</td>
</tr>
<tr>
<td>4.5. Prognostic evaluation*</td>
<td>c) Evaluating the likely course of the performance. (Ortega, 2005)</td>
</tr>
<tr>
<td></td>
<td>e.g.) G6−2: At this point when I was writing in English, I was thinking about mostly whether I would be able to translate it or not.</td>
</tr>
</tbody>
</table>

* Strategy types that Ortega (2005) created

** Strategy types that the present study added to Ortega’s (2005) classifications

a) Definitions from O’Malley and Chamot (1990)

b) Definitions from Oxford (1990)

c) Definitions from Ortega (2005)

d) Definitions that were created for the present study
<table>
<thead>
<tr>
<th>Strategy categories</th>
<th>Definitions &amp; Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Writing for retrieval*</td>
<td>c) Writing out to retrieve one’s language knowledge. e.g.) I wasn’t sure what is the te-form of <em>iku</em>. I was writing it out so the right form would come up somehow.</td>
</tr>
<tr>
<td>1.2. Elaboration</td>
<td>a) Relating new information to prior knowledge; relating different parts of new information to each other; making meaningful personal associations to information presented (p. 138).</td>
</tr>
<tr>
<td>2.1 Writing for later recall*</td>
<td>c) Writing down so one can remember better. e.g.) G5−3: I wrote that down, that helps reinforce remembering it.</td>
</tr>
<tr>
<td>2.2. Rehearsing</td>
<td>a) Imitating a language model, including overt practice and silent rehearsal (Repetition, p. 119). Rehearsing includes mental rehearsal (rehearsing in one’s head), subarticular rehearsal (rehearsing by whispering), reading rehearsal (rehearsing by reading one’s notes), and selective rehearsal (rehearsing parts of the task) (Ortega, 2005). e.g.) G5−2: I decided to go through it in my head once.</td>
</tr>
<tr>
<td>3.1. Highlight and postpone</td>
<td>b) <em>Using a variety of emphasis techniques</em> (such as underlining, starring, or color-coding) to focus on important information in a passage (Highlighting, p. 47). e.g.) I wrote <em>akeru</em>, but I wasn’t sure if it was the right verb. I underlined it, so I would come back later.</td>
</tr>
<tr>
<td>3.2. Make up</td>
<td>b) <em>Making up new words</em> to communicate the desired idea, such as <em>paper-holder</em> for <em>notebook</em> (Compensation strategies, Coining words, p. 50). e.g.) I wasn’t sure what the Japanese word for the power cord, so I said <em>conpyuta no himo</em> [a string for a computer].</td>
</tr>
<tr>
<td>4. Avoidance</td>
<td>b) <em>Partially or totally avoiding communication</em> when difficulties are anticipated (Compensation strategies, Avoiding communication partially or totally p. 50).</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>e.g.) G6−2: I thought about how I wished to describe their emotions in the last panel, cause it looked like they all had different expressions, and I thought that would be hard to explain in Japanese. I wasn’t sure how I would connect the sentences, so I decided to pretend that the other student wasn’t there, to make it simpler.</td>
<td></td>
</tr>
<tr>
<td>5.1. Approximating</td>
<td>b) <em>Altering the message</em> by omitting some items of information, making ideas simpler or less precise, or saying something slightly different that means almost the same thing, such as <em>pencil</em> for <em>pen</em> (Compensation strategies, Adjusting or approximating the message p. 50).</td>
</tr>
<tr>
<td>e.g.) G5−1: So after, I didn’t know if I could use <em>tsukeru</em>, so I used <em>tsukau</em>. So, instead of “turning on,” I changed my approach to “to use.” Basically telling the story in a different angle, I guess.</td>
<td></td>
</tr>
<tr>
<td>5.2. Circumlocution and synonyms</td>
<td>b) <em>Getting the meaning across by describing the concept (circumlocution) or using a word that means the same thing (synonym)</em>; for example, “what you use to wash dishes with” as a description for <em>dishrag</em> (Compensation strategies, Using a Circumlocution or Synonym, p. 51).</td>
</tr>
<tr>
<td>e.g.) I didn’t know how to call the power cable, so I decided to say <em>conpyuta wo tsukeru tame no kodo</em> [the cord to turn on the computer].</td>
<td></td>
</tr>
<tr>
<td>5.3. Switching to the mother tongue**</td>
<td>b) Using the mother tongue for an expression without translating it…. This strategy may also include adding word endings from the new language onto words from the mother tongue (p. 50).</td>
</tr>
<tr>
<td>e.g.) G3−3: [I: When you had the vocabulary problem, what did you decide to do?] Just substituting in English, unfortunately.</td>
<td></td>
</tr>
</tbody>
</table>
Table 9. Continued

<table>
<thead>
<tr>
<th>Strategy Type</th>
<th>Example</th>
</tr>
</thead>
</table>
| 5.4. Other lexical compensation | d) Other ways of compensating for lack of vocabulary.  
  e.g.) G1−3: Trying to say it with words that I do know. |
| 6. Translating | a) Rendering ideas from one language to another in a relatively verbatim manner (Translation, p. 138).  
  e.g.) G5−2: I actually went through it and translate it in my head. |
| 7. Across-language analysis | b) Comparing elements (sounds, vocabulary, grammar) of the new language with elements of one’s own language to determine similarities and differences (Analyzing contrastively, p. 46).  
  e.g.) Since he looked like a teacher, I thought I should use the honorific form to describe him. In English, we don’t have the honorific form, but in Japanese you need to use it. |
| 8. Outlining/summarizing | a) Making a mental or written summary of language and information presented in a task (Summarizing, p. 138).  
  e.g.) G5−3: I just went kind of frame by frame with all the pictures and wrote a few ideas and some vocabulary, kind of something down for each one just to kind of have some ideas. |

* Strategy types that Ortega (2005) created  
** Strategy types that the present study added to Ortega’s (2005) classifications  
  a) Definitions from O’Malley and Chamot (1990)  
  b) Definitions from Oxford (1990)  
  c) Definitions from Ortega (2005)  
  d) Definitions that were created for the present study
Table 10. Definitions and examples of social/affective strategies

<table>
<thead>
<tr>
<th>Strategy categories</th>
<th>Definitions &amp; Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lowering anxiety</td>
<td>b) Using the technique of <em>alternately tensing and relaxing</em> all of the major muscle groups in the body, … in order to relax; or the technique of <em>breathing deeply</em> from the diaphragm; or the technique of <em>meditating by focusing</em> on a mental image or sound (Using progressive relaxation, deep breathing, or meditation, p. 143).</td>
</tr>
<tr>
<td></td>
<td>b) <em>Using laughter to relax</em> (Using laughter, p. 143).</td>
</tr>
<tr>
<td></td>
<td>e.g.) G6−1: I knew I could explain it, and I was just focused on trying to not be nervous.</td>
</tr>
<tr>
<td>2. Encouraging oneself</td>
<td>b) <em>Saying or writing positive statements</em> to oneself in order to feel more confident in learning the new language (Making positive statements, p. 143).</td>
</tr>
<tr>
<td></td>
<td>b) <em>Pushing oneself to take risks</em> in a language learning situation, even though there is a chance of making a mistake or looking foolish. Risks must be tempered with good judgment (Taking risks wisely, p. 144).</td>
</tr>
<tr>
<td></td>
<td>b) <em>Giving oneself a valuable reward</em> for a particularly good performance in the new language (Rewarding yourself, p. 144).</td>
</tr>
<tr>
<td></td>
<td>e.g.) G5−1: I was probably thinking, “OK, last one, just think.”</td>
</tr>
<tr>
<td>3. Empathizing with the listener</td>
<td>c) Thinking of the listener’s needs (p. 98).</td>
</tr>
<tr>
<td></td>
<td>e.g.) G6−1: The fifth picture, I believe it was, trying to explain the fact that the power cord was unplugged to do it without taking a long time to be brief and clear so that it would be easy to understand for somebody who hasn’t seen the pictures.</td>
</tr>
</tbody>
</table>

* Strategy types that Ortega (2005) created

** Strategy types that the present study added to Ortega’s (2005) classifications

a) Definitions from O’Malley and Chamot (1990)

b) Definitions from Oxford (1990)

c) Definitions from Ortega (2005)

d) Definitions that were created for the present study
3.6.2.2 Interrater reliability to evaluate strategies

To establish interrater reliability, a second rater, who was different from the rater who coded the participants’ narrative stories, was assigned. She received the strategy charts (Tables 8–10, pp. 77–83) and was instructed how to code the strategies based on the strategy categories. Then, she analyzed three interviews together with the researcher for training purposes. After this first session, the second rater received a set of six interviews, (25% of the interview data), one interview each from the six experimental groups. She and the researcher individually analyzed them and her coding was compared to the researchers’ coding. The interrater reliability was 80.65%. Inconsistencies in the strategy coding were resolved by discussion until the second rater and the researcher reached 100% agreement. The rest of the interviews were analyzed by the researcher alone, following the coding scheme determined in conjunction with the second rater.

3.7 Analysis

3.7.1 Analysis of oral production

This study employed a between-subjects design, and each participant performed story-telling tasks under one planning condition. Evaluation of the participants’ oral production was composed of three statistical analyses. The first analysis explored the possible effects of time pressure on fluency in the L2 learners’ oral production (research question 1). The primary purpose of the second quantitative analysis was to investigate how different planning conditions interacted with the fluency, complexity, and accuracy of the L2 learners’ resulting oral production (research questions 2.1–2.3). The third quantitative analysis examined possible trade-off effects by means of correlations (research question 2.4). Statistical methods that were employed for each analysis are explained in the following sections.
3.7.1.1 Analysis for research question 1

Research question 1 asks whether time pressure affects L2 oral production, particularly fluency. Without time pressure, the participants may engage in different amounts of on-line planning during the task, which may affect the results, as Yuan and Ellis (2003) suggested. However, by the same token, it is possible that time pressure pushes the learners to speak faster, resulting in greater fluency. The impact of time pressure on fluency, in the absence of on-line planning, was determined in this first analysis so that the second analysis would compare specifically the effects of each planning condition. Thus, the first analysis was a necessary prior stage that underpins the second quantitative analysis.

The following four groups were compared for the first analysis: G1 (control; no pre-task, no time pressure), G2 (no pre-task planning, time pressure), G3 (pre-task planning, no time pressure), G4 (pre-task planning, time pressure). The independent variables were time pressure and pre-task planning, both of which had two levels (with and without time pressure and with and without pre-task planning). The dependent variables were fluency, complexity, and accuracy. A two-way MANCOVA was performed on the measures that were selected for each dependent variable, based on the results of factor analysis. SPOT scores were used as a covariate to analyze the participants’ oral production. The alpha level was set at .05.

3.7.1.2 Analysis for research question 2

Research question 2.1–3 addresses the effects of different types of planning on fluency, complexity, and accuracy of the L2 oral production. If the first analysis for research question 1 described above (section 3.7.1.1) shows that time pressure affected the participants’ fluency, the speech data from G2 and G4, who performed a task under time pressure, will be excluded from this second analysis. If, however, the first analysis shows that time pressure had no influence on L2 learners’ oral production, including
fluency, the data from G1 and G2 would be combined into a single no pre-task planning group (NPG). The same would be true for G3 and G4 groups: They would be combined as a pre-task planning group (PG). In this case, the following four groups would be compared: NPG (G1 and G2 combined), PG (G3 and G4 combined), G5, and G6. The experimental groups are displayed in Table 11.

Table 11. Experimental groups to be compared for research question 2

<table>
<thead>
<tr>
<th>Experimental groups</th>
<th>Pre-task planning</th>
<th>Prompt for on-line planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPG (= G1 &amp; G2)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>PG (= G3 &amp; G4)</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>G5</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>G6</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

In either case, the independent variables were pre-task planning and on-line planning and the dependent variables were fluency, complexity, and accuracy. A MANCOVA, again using the scores of the SPOT as a covariate, was performed on the measures for fluency, complexity, and accuracy to analyze the data. The alpha level was set at .05.

The third quantitative analysis was conducted to answer research question 2.4, which asks whether there are trade-off effects among fluency, complexity, and accuracy. Correlations among the measures for these three aspects were evaluated pair-wise within each group and also for all the participants regardless of planning condition to identify if
there were any universal relationships among these three aspects of L2 oral production. The data from the same groups that were involved in the second analysis were examined for this third analysis.

3.7.2 Analysis of learner strategies for research question 3

The analysis of learner strategies was conducted to explore the overall processes of task planning and to identify strategies that L2 learners used during planning (research question 3). In this study, focus was on which strategies the participants used, not how many times they used each strategy.

To determine whether there were any differences in the selections of strategy type depending on the planning condition that participants were provided, the strategies that each participant used were examined, and the data were compared between the strategy use during pre-task planning and during task performance. The data during task performance was further divided into two based on whether the participants were encouraged to plan on-line or not.
CHAPTER 4: RESULTS OF L2 ORAL PRODUCTION

4.1 Introduction

This chapter presents the results of the quantitative analyses on L2 learners’ oral production in this study. It starts with reporting the results of the factor analysis that was performed on L2 measures of fluency, complexity, and accuracy in the participants’ narrative stories to determine the most appropriate dependent variables for this study. Correlations among these measures were also examined, and the results are reported. Based on the findings for the factor analysis and correlations, five measures were employed as dependent variables for this study: one fluency measure, two complexity measures, and two accuracy measures. Using these measures, a two-way MANCOVA was performed on the data to examine the effects of time pressure on L2 learners’ oral performance, and the results are given to answer research question 1. In response to research question 2, the results of another two-way MANCOVA as well as those of correlations that explored the effects of planning on L2 learners’ oral performance are presented. The qualitative analyses in response to research question 3 are presented in the next chapter.

4.2 Selection of measures for speech data

To determine which dependent variables were most appropriate to examine participants’ oral production, a factor analysis was performed on the 12 measures that were described in chapter 3. The data from all 143 participants were subjected to a principal component analysis, and the extracted factors were rotated by the Varimax rotation method. Table 12 presents part of the factor structure that is relevant to the purpose of this analysis, and the complete factor structure is displayed in Appendix F.
Table 12. Rotated component matrix

<table>
<thead>
<tr>
<th></th>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td></td>
<td>.940</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pruned speech rate</td>
<td></td>
<td>.938</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pauses</td>
<td></td>
<td>-.857</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses/T-unit</td>
<td></td>
<td>.848</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words/T-unit</td>
<td></td>
<td>.861</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td></td>
<td>.110</td>
<td></td>
<td>.869</td>
<td></td>
</tr>
<tr>
<td>Modified TTR</td>
<td></td>
<td>.744</td>
<td></td>
<td>.087</td>
<td></td>
</tr>
<tr>
<td>Error-free clauses</td>
<td></td>
<td>.546</td>
<td>.546</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error rate</td>
<td></td>
<td>-.545</td>
<td></td>
<td>-.651</td>
<td>.400</td>
</tr>
<tr>
<td>Pruned error rate</td>
<td></td>
<td>-.494</td>
<td></td>
<td>-.641</td>
<td>.475</td>
</tr>
<tr>
<td>Correct particles</td>
<td></td>
<td></td>
<td></td>
<td>.901</td>
<td></td>
</tr>
<tr>
<td>Correct verb forms</td>
<td></td>
<td>.482</td>
<td>.220</td>
<td></td>
<td>-.453</td>
</tr>
</tbody>
</table>

Of the four factors that are displayed in Table 12, three were extracted with eigenvalues greater than 1. The fourth factor (eigenvalue of .96) was added to further investigate the inter-relationship of the measures. In examining the factor loadings, two measures for syntactic complexity (number of clauses per T-unit and number of words per T-unit) had high loadings on the first factor. These results indicate that this factor concerns grammatical complexity. Modified type-token ratio (TTR) loaded highly on the first factor as well, although syntactic complexity and lexical complexity do not necessarily correspond to each other. Moderate loadings were observed on four of the five accuracy measures (positive loadings of .546 and .482 for error free clauses and correct verb forms and negative loadings of -.545 and -.494 for error rate and pruned error rate), and it appears that syntactic complexity has some link with grammatical accuracy.
The second factor can be clearly interpreted as fluency, defined by three fluency measures, loading highly on both unpruned and pruned speech rates and loading negatively on the number of pauses. The rest of the measures show minor loadings for the second factor, ranging from \(-0.146\) to \(0.198\).

The third factor seems to be an indicator for accuracy, loading highly on four of the five accuracy measures, particularly the correct particle with a factor loading of \(0.901\). The correct particle measure is a specific accuracy measure, and the third factor loaded on all the global accuracy measures as well: \(0.546\) for the error free clauses, \(-0.651\) and \(-0.641\) for the error rate and the pruned error rate. The only measure of accuracy that did not have a significant loading on the third factor is the percentage of correct verb forms, which is the other specific accuracy measure. It seems that accurate use of particles does not relate to the performance on producing correct verb forms, and vice versa. No significant loadings were obtained on other measures for this factor.

As explained earlier, the fourth factor was added to the results although its eigenvalue did not reach 1. Inclusion of the fourth factor was intended to find the connection of TTR to other measures, especially to modified TTR because the loading on TTR remained low for the first three factors, even when the modified TTR showed a high loading on the first factor. By including the fourth factor, it is clear that TTR does not have strong connection with any of the measures that were examined by the factor analysis. Considering the low loading of modified TTR on the fourth factor, these two measures seem to be independent although both measures are supposed to assess the variety of vocabulary in language production. The fourth factor also had some loadings on three of the accuracy measures: error rate, pruned error rate, and the percentage of correct verb forms. It appears that these measures have connection with variety of vocabulary to some degree. The remaining measures loaded considerably low on the fourth factor, from \(-0.272\) to \(0.109\).
Turning to the question of what each measure concerns, all of the three fluency measures loaded significantly on the second factor, but not on any other factors. None of the measures for complexity or accuracy loaded on the second factor and, therefore, these measures clearly indicate fluency in the data.

As to complexity measures, two measures each for syntactic complexity and lexical complexity were tested in the factor analysis. Both syntactic measures (clauses per T-unit and words per T-unit) had high loadings on the first factor only. Four of the five accuracy measures also loaded on the first factor, which seems to suggest that the first factor indicates syntactic accuracy as well as syntactic complexity. Thus, both clauses per T-unit and words per T-unit may be related to syntactic accuracy.

While two syntactic complexity measures showed similar loading patterns to each other, the lexical complexity measures loaded on different factors. TTR loaded highly on the fourth factor and had minor loadings on the other three factors. In contrast, modified TTR loaded on the first factor only, and not on the rest of the factors, including the fourth factor on which TTR showed a high loading. As stated earlier, the first factor appears to be an indicator of syntactic complexity and accuracy, which should be independent from lexical complexity.

Finally, the factor analysis evaluated five accuracy measures: three global accuracy measures and two specific accuracy measures. The correct use of particles loaded highly on the third factor, and did not load on any other factors. All three global accuracy measures also loaded on the third factor, although these measures loaded on other factors, too. Error-free clauses loaded on the first factor in addition to the third factor, which indicates that this measure can be associated with syntactic complexity. Both the unpruned and pruned error rates loaded moderately on three factors: the first, third, and fourth factors. Therefore, these measures may be related to both syntactic and lexical complexity. The percentage of correct verb forms was the only accuracy measure that had a low loading on the third factor. Instead of the third factor, it loaded on the first
and fourth factors, which may suggest that this measure has some connection with syntactic and lexical complexity rather than with accuracy.

To further analyze the relation among dependent variables and determine the most appropriate measures for each of fluency, complexity, and accuracy, Pearson’s correlation within each group of measures was performed in addition to the factor analysis. The complete correlation matrix is displayed in Appendix G. Table 13 shows the correlation matrix among fluency measures: speech rate, pruned speech rate, and the number of pauses.

Table 13. Correlations among fluency measures

<table>
<thead>
<tr>
<th></th>
<th>Speech rate</th>
<th>Pruned speech rate</th>
<th>Pauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pruned speech rate</td>
<td>.973**</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Pauses</td>
<td>-.720**</td>
<td>-.703**</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

As Table 13 shows, speech rate was highly correlated with pruned speech rate at .973, which suggests that participants who generally spoke faster maintained high speech rate even after removing repeated, rephrased, and reformulated part of their narratives. Strong negative correlations were found for the number of pauses with both unpruned and pruned speech rates as well. This result indicates that participants with high speech rates had fewer pauses and vice versa.

In the case of complexity, correlations were examined among the number of clauses per T-unit (syntactic complexity), the number of words per T-unit (syntactic complexity), TTR (lexical complexity), and modified TTR (lexical complexity). The
correlations among these measures were not straightforward. The results are displayed in Table 14.

Table 14. Correlations among complexity measures

<table>
<thead>
<tr>
<th></th>
<th>Clauses/T</th>
<th>Words/T</th>
<th>TTR</th>
<th>Modified TTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clauses/T</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words/T</td>
<td>.855**</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>.028</td>
<td>-.147</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Modified TTR</td>
<td>.496**</td>
<td>.485**</td>
<td>.108</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Regarding syntactic complexity, the number of clauses per T-unit was highly correlated with the number of words per T-unit (.855). However, two measures for lexical complexity had a low correlation at .108 between them, and did not reach a significant level. In examining the correlations of each lexical measure with syntactic complexity measures, TTR did not correlate with either of the complexity measures, whereas modified TTR showed moderate correlations with both of them. In addition to the results of the factor analysis, these results further suggest that modified TTR may be related to grammatical aspects of language rather than vocabulary.

In terms of accuracy measures, correlations among five dependent variables were examined: the percentage of error-free clauses (global accuracy), error rate (global accuracy), pruned error rate (global accuracy), the percentage of correct use of particles (specific accuracy), and the percentage of correct use of verb forms (specific accuracy).
The results show that three of the global accuracy measures had high correlations with each other. The correlation between the error rate and the pruned error rate was especially high at .965, and the error-free clauses negatively correlated with these two measures at a high level as well: −.750 with the error rate and −.734 with the pruned error rate. Concerning the specific accuracy measures, the correct particles and correct verb forms showed no correlation with each other, which suggests that accuracy of particles has little to do with accuracy of verb forms. In regard to the correlations between global measures and specific measures, correct verb form was moderately correlated with all three global measures, whereas correct particles showed low to moderate correlations with them. The correlation matrix is displayed in Table 15.

Table 15. Correlations among accuracy measures

<table>
<thead>
<tr>
<th></th>
<th>Error-free</th>
<th>Error rate</th>
<th>Pruned error rate</th>
<th>Correct particle</th>
<th>Correct verb form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error-free</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error rate</td>
<td>−.750**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pruned error rate</td>
<td>−.734**</td>
<td>.965**</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct particle</td>
<td>.325**</td>
<td>−.438**</td>
<td>−.439**</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Correct verb form</td>
<td>.544**</td>
<td>−.598**</td>
<td>−.556**</td>
<td>.098</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Observing the loading patterns that were shown by the factor analysis and the correlation patterns, the following five measures seemed to be the best predictor of fluency, complexity, and accuracy of L2 oral production and, therefore, were employed to analyze the participants’ narrative stories for this study:
Fluency: Number of moras per minute
Complexity: Number of clauses per T-unit (syntactic complexity)
            Type-token ratio (lexical complexity)
Accuracy:  Percentage of error-free clauses (global accuracy)
            Percentage of correct use of particles (specific accuracy)

In the case of fluency, all three of the fluency measures were found to be highly loaded on the second factor and had strong correlations to each other, indicating that they are reflecting the same trait in the data. Among the three fluency measures, the unpruned speech rate exhibited the highest loading on the second factor, although the pruned speech rate loaded nearly as highly, and had stronger correlations with the other two fluency measures. Thus, the unpruned speech rate (i.e., the number of moras per minute) was considered to be the most appropriate representation of learner fluency of the three fluency measures.

Regarding syntactic complexity, both of the measures had high loadings on the first factor, and they are highly correlated with each other, indicating that these two measures are strongly related. Therefore, the number of clauses per T-unit, which has also been commonly used in past planning studies (Kawahuchi, 2005a, 2005b; Yuan & Ellis, 2003), was selected as a syntactic complexity measure for this study.

For lexical complexity, type-token ratio (TTR) loaded highly on the fourth factor, which appears to indicate lexical complexity as well as lexical accuracy. On the other hand, modified TTR loaded highly on the first factor, which seems to be an indicator for syntactic complexity and accuracy. The results of correlations further suggest that modified TTR is related to syntactic complexity, whereas TTR is not. Thus, TTR was concluded to be an optimal measure for lexical complexity in this study, preferable to modified TTR.
As for accuracy measures, one measure each was selected from the global and specific measures. For the global measures, the percentage of error-free clauses was selected because it showed a similar loading pattern to the other two measures for the first and third factors and because it had high correlations with both of them. The percentage of error-free clauses has been used in many planning studies (Mehnert, 1998; Skehan & Foster, 1997, 2005; Yuan & Ellis, 2003) and, therefore, it can be considered a reasonable choice because it will allow for comparisons between the results of this study and those of past studies.

For the specific accuracy measure, the percentage of correct use of particles and the percentage of correct verb forms were examined. The percentage of correct use of particles loaded highly on the third factor, which seems to specifically reflect grammatical accuracy because this factor also loaded highly on the three global accuracy measures, but not on either fluency or complexity measures. The percentage of correct verb forms, in contrast, had a minor loading on the third factor. Therefore, the percentage of correct use of particles was selected for specific accuracy measures.

These five measures were used as dependent variables for this study and the statistical analyses were performed with these variables. The following sections present the results of these analyses.

4.3 Effects of task conditions

Using the five measures selected in the last section, means and standard deviations were calculated by group for the six different task conditions. The results show that the participants in G4, who had an opportunity to plan prior to the task performance and had a time limit during the task, spoke the fastest among the six planning groups, generating a mean of 216.28 moras per minute on average. The participants in G5, who planned on-line, but had no pre-task planning opportunity and no time limit to perform a task, spoke the slowest at 139.61 moras per minute.
Regarding syntactic complexity, the narrative stories by G6, who planned both before and during the task, contained the most syntactically complex sentences, averaging 2.07 clauses per T-unit, whereas those by G3, who had a pre-task planning opportunity only and performed the task without a time limit, were the least syntactically complex with 1.78 clauses per T-unit.

The results of type-token ratio (TTR) revealed an interesting picture. G1 and G2, neither of whom had planning opportunities, produced narrative stories with the highest TTR (0.53). In contrast, G6, who planned both before and during the task, produced the stories with the least variety of vocabulary (TTR = 0.43). G3, who planned before the task, produced stories with the second lowest variety in vocabulary (TTR = 0.46).

As regards accuracy, different results were obtained for the global and specific measures. Concerning the global accuracy measure, defined as the percentage of error-free clauses, G6 produced the most accurate narrative stories (39.61%), and G2’s were the least accurate (28.28%). However, in the case of the specific measure (the percentage of correct use of particles), G2 was the most accurate at 89.98% accuracy rate. G3 had the lowest mean accuracy rate at 86.12%. Table 16 presents the descriptive statistics.
The data were submitted to statistical analyses first to analyze the effects of time pressure for research question 1, and then those of planning types for research question 2. The results are presented in the next two sections.

4.3.1 Time pressure

The first research question is as follows:

1. How does time pressure affect the L2 learners’ oral production in a narrative task?

For this research question, the data from G1, G2, G3, and G4 were used. G1 and G3 were combined to form the No Time Pressure group (NTPG), and G2 and G4 were
combined to form the Time Pressure group (TPG). For an additional comparison, G1 and G2 were also combined to form the No Pre-Task Planning group (NPG) and G3 and G4 as the Pre-Task Planning group (PG). Table 17 displays a summary of the task conditions for each group.

<table>
<thead>
<tr>
<th></th>
<th>Time pressure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>G1 (n = 24)</td>
<td>G2 (n = 25)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>G3 (n = 24)</td>
</tr>
<tr>
<td>Total</td>
<td>NTPG (n = 48)</td>
<td>TPG (n = 50)</td>
</tr>
</tbody>
</table>

As a provisional step, the time spent to complete the task by each group was compared to confirm that the participants in the TPG indeed performed within the assigned time limit and that the NTPG group needed time longer than the time limit to complete the task. Table 18 shows how much time each group spent on the task.
Table 18. Descriptive statistics for time in seconds that was spent to complete the task

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SD)</th>
<th>Group</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTPG (n = 48)</td>
<td></td>
<td>TPG (n = 50)</td>
<td></td>
</tr>
<tr>
<td>G1 (n = 24)</td>
<td>100.34 (51.81)</td>
<td>G2 (n = 25)</td>
<td>78.07 (23.80)</td>
</tr>
<tr>
<td>G3 (n = 24)</td>
<td>132.29 (50.54)</td>
<td>G4 (n = 25)</td>
<td>84.62 (25.81)</td>
</tr>
<tr>
<td>Total</td>
<td>116.31 (53.14)</td>
<td>Total</td>
<td>81.34 (24.79)</td>
</tr>
</tbody>
</table>

Overall, the NTPG spent 116.31 seconds on average, which was longer than the time limit of 90 seconds. Individually, G1 spent 100.34 seconds and G3 spent 132.29 seconds; therefore, the task appears to require more than 90 seconds to complete without time pressure. On the other hand, the TPG spent 81.34 seconds on average to perform the task, which is within the time limit. G2 spent 78.07 seconds, whereas G4 spent 84.62 seconds, and neither group exceeded the time limit. Considering these results, it seems clear that the participants in the TPG felt compelled to keep their performance within the time limit.

As explained earlier, there were five measures to analyze the participants’ narrative stories: speech rate (fluency), number of clauses per T-unit (syntactic complexity), type-token ratio (lexical complexity), percentage of error-free clauses (global accuracy), and percentage of correct use of particles (specific accuracy). The mean and the adjusted mean with covariate (SPOT scores) for each measure by group is presented in Tables 19–23. Since the focus of this analysis is to investigate the effects of time pressure on L2 learners’ oral production, the results of the NPG and PG are for reference only, and are not discussed in this section.
Comparing the speech rate between the NTPG and the TPG, the NTPG produced 169.46 moras per minute, whereas the TP group produced 199.15 moras per minute, which means that the TPG outperformed the NTPG by approximately 30 moras per minute. Table 19 displays the results for speech rate.

Table 19. Descriptive statistics for speech rate

<table>
<thead>
<tr>
<th></th>
<th>Time pressure</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>−</td>
<td>176.64 (57.58)</td>
<td>182.03 (49.02)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>162.27 (34.59)</td>
<td>216.28 (49.91)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td></td>
<td>169.46 (47.55)</td>
<td>199.15 (51.93)</td>
</tr>
<tr>
<td>Adjusted mean</td>
<td></td>
<td>170.88</td>
<td>198.64</td>
</tr>
</tbody>
</table>

As for syntactic complexity, the TPG had 1.99 clauses per T-unit and outperformed the NTPG, which produced 1.86 clauses per T-unit (Table 20). Regarding lexical variety, the TPG outperformed the NTPG in terms of the variety of vocabulary as well. The TTR for the TPG was .51, whereas that of the NTPG was .49 (Table 21).
Table 20. Descriptive statistics for the number of clauses per T-unit

<table>
<thead>
<tr>
<th></th>
<th>Time pressure</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pre-task planning −</td>
<td>Mean (SD)</td>
<td>1.94 (.45)</td>
<td>2.03 (.78)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Mean (SD)</td>
<td>1.78 (.38)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>1.86 (.42)</td>
<td>1.99 (.69)</td>
</tr>
<tr>
<td></td>
<td>Adjusted mean</td>
<td>1.87</td>
<td>1.98</td>
</tr>
</tbody>
</table>

Table 21. Descriptive statistics for type-token ratio

<table>
<thead>
<tr>
<th></th>
<th>Time pressure</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pre-task planning −</td>
<td>Mean (SD)</td>
<td>.53 (.08)</td>
<td>.53 (.07)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Mean (SD)</td>
<td>.46 (.08)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>.49 (.09)</td>
<td>.51 (.07)</td>
</tr>
<tr>
<td></td>
<td>Adjusted mean</td>
<td>.49</td>
<td>.51</td>
</tr>
</tbody>
</table>

The results for the global measure revealed that the NTPG and the TPG obtained almost the same accuracy rate (NTPG = 33.74%; TPG = 33.27%). Turning to the specific accuracy measure, the results show that the accuracy rate for the TPG was slightly higher
than that of the NTPG (NTPG = 87.42%; TPG = 88.53%). Tables 22 and 23 show the results for the global and specific accuracy.

Table 22. Descriptive statistics for percentage of error-free clauses

<table>
<thead>
<tr>
<th></th>
<th>Time pressure</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>−</td>
<td>34.26 (13.79)</td>
<td>28.38 (16.51)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>33.21 (17.80)</td>
<td>38.16 (18.10)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>33.74 (15.76)</td>
<td>33.27 (17.85)</td>
</tr>
<tr>
<td>Adjusted mean</td>
<td></td>
<td>33.96</td>
<td>33.05</td>
</tr>
</tbody>
</table>

Table 23. Descriptive statistics for percentage of correct use of particles

<table>
<thead>
<tr>
<th></th>
<th>Time pressure</th>
<th>Mean (SD)</th>
<th>Adjusted mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>−</td>
<td>88.72 (6.64)</td>
<td>89.98 (8.49)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>86.12 (9.73)</td>
<td>87.08 (7.71)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>87.42 (8.35)</td>
<td>88.53 (8.16)</td>
</tr>
<tr>
<td>Adjusted mean</td>
<td></td>
<td>87.47</td>
<td>88.48</td>
</tr>
</tbody>
</table>
To investigate the effects of time pressure on L2 learners’ oral production, these results were submitted to a two-way MANCOVA using time limit and pre-task planning as independent variables, five measures of the features of the narrative stories as dependent variables, and SPOT scores as covariates. The alpha level was set at .05. Assumptions for a MANCOVA were evaluated, and it was found that they were not violated. Table 24 displays the results of Wilks’ lambda for independent variables. As mentioned earlier, the results of pre-task planning are for reference only for this analysis and are not discussed.

Table 24. Wilks’ lambda for a two-way MANCOVA for time pressure and pre-task planning effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time pressure</td>
<td>.851</td>
<td>3.105</td>
<td>5</td>
<td>89</td>
<td>.012*</td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>.789</td>
<td>4.760</td>
<td>5</td>
<td>89</td>
<td>.001**</td>
</tr>
<tr>
<td>Time pressure × Pre-task planning</td>
<td>.898</td>
<td>2.030</td>
<td>5</td>
<td>89</td>
<td>.082</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Wilks’ lambda revealed significant main effects for time limit ($F = 3.105, p = .012$). However, the interaction between time limit and pre-task planning did not reach a significant level ($F = 2.030, p = .082$) and, therefore, any significant results that would be obtained for the interaction by a MANCOVA will be disregarded.

Since the results of the MANCOVA showed that time pressure was significant, the follow-up ANCOVAs were performed on five dependent variables: speech rate,
number of clauses per T-unit, type-token ratio, percentage of error-free clauses, and percentage of correct use of particles. The results are displayed by each dependent variable in Tables 25–29. The covariate was significant for all the dependent variables, except for TTR.

Table 25 shows the results for fluency. It was found that there was a significant difference for speech rate between the NTPG and TPG ($F = 10.361$, $p = .002$). The speech rate for the TPG was higher than that for the NTPG (cf. Table 19), which suggests that L2 learners speak faster when they have a time limit to perform a task. For fluency, the interaction between time pressure and pre-task planning was also significant. However, as observed earlier in Table 24, the Wilks’ lambda analyses revealed that the overall interaction effect between time pressure and pre-task planning was not significant, and therefore, the significant interaction that was found in fluency is not considered in this study.

Table 25. Follow-up ANCOVA for time pressure and pre-task planning effects on speech rate

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time pressure</td>
<td>20083.868</td>
<td>10.361</td>
<td>.002**</td>
</tr>
<tr>
<td>Pre-task</td>
<td>2171.309</td>
<td>1.120</td>
<td>.293</td>
</tr>
<tr>
<td>Time pressure × Pre-task</td>
<td>12963.626</td>
<td>6.688</td>
<td>.011*</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
The results for syntactic complexity are displayed in Table 26. The results revealed no significant difference between the NTPG and the TPG.

Table 26. Follow-up ANCOVA for time pressure and pre-task planning effects on the number of clauses per T-unit

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time pressure</td>
<td>.284</td>
<td>1.216</td>
<td>.273</td>
</tr>
<tr>
<td>Pre-task</td>
<td>.419</td>
<td>1.797</td>
<td>.183</td>
</tr>
<tr>
<td>Time pressure × Pre-task</td>
<td>.010</td>
<td>.043</td>
<td>.836</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Table 27 shows the results for lexical complexity. The results, again, show no significant difference between the NTPG and the TPG for time pressure.

Table 27. Follow-up ANCOVA for time pressure and pre-task planning effects on TTR

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time pressure</td>
<td>.008</td>
<td>1.411</td>
<td>.238</td>
</tr>
<tr>
<td>Pre-task</td>
<td>.067</td>
<td>11.285</td>
<td>.001**</td>
</tr>
<tr>
<td>Time pressure × Pre-task</td>
<td>.003</td>
<td>.519</td>
<td>.473</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
Regarding global accuracy, the results of the follow-up ANCOVA are shown in Table 28. No significant differences were found for the percentage of error-free clauses.

Table 28. Follow-up ANCOVA for time pressure and pre-task planning effects on the percentage of error-free clauses

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time pressure</td>
<td>20.13</td>
<td>0.099</td>
<td>0.753</td>
</tr>
<tr>
<td>Pre-task</td>
<td>420.98</td>
<td>2.078</td>
<td>0.153</td>
</tr>
<tr>
<td>Time pressure × Pre-task</td>
<td>582.62</td>
<td>2.876</td>
<td>0.093</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Finally, as to specific accuracy, no significant difference was found between the NTPG and the TPG. The results are displayed in Table 29.

Table 29. Follow-up ANCOVA for time pressure and pre-task planning effects on the percentage of correct use of particles

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time pressure</td>
<td>24.91</td>
<td>0.387</td>
<td>0.535</td>
</tr>
<tr>
<td>Pre-task</td>
<td>192.16</td>
<td>2.984</td>
<td>0.087</td>
</tr>
<tr>
<td>Time pressure × Pre-task</td>
<td>1.84</td>
<td>0.029</td>
<td>0.866</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
The results of the follow-up ANCOVAs revealed that there was a significant difference in speech rate between the NTPG and TPG. However, the comparison between these two groups did not show any significant differences in the rest of the dependent variables, which means that time pressure improved specifically the fluency in the participants’ task performance, but did not affect either complexity or accuracy.

4.3.2 Task planning

The second research question is concerned with whether there are any differences in the way that pre-task planning and on-line planning affect the L2 learners’ oral production in a narrative task, controlling for pre-existing Japanese proficiency differences. In examining this question, the narrative stories that were produced by the participants in G1, G3, G5, and G6 were compared. For this analysis, G1 and G5 were combined as a No Pre-Task Planning group (NPG2, to be distinguished from the NPG in the prior section), whereas G3 and G6 were combined to be a Pre-Task Planning group (PG2). G1 was also merged with G3 as the No On-line Planning group (NOLP), and G5 and G6 are the On-line Planning group (OLP). The grouping is summarized in Table 30.

Table 30. Task conditions for G1, G3, G5, and G6

<table>
<thead>
<tr>
<th></th>
<th>On-line planning</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>G1 (n = 24)</td>
<td>G5 (n = 24)</td>
</tr>
<tr>
<td></td>
<td>G3 (n = 24)</td>
<td>G6 (n = 21)</td>
</tr>
<tr>
<td>Total</td>
<td>NOLP (n = 48)</td>
<td>OLP (n = 45)</td>
</tr>
</tbody>
</table>
Participants’ narrative task performance was analyzed for five dependent variables by group: speech rate, number of clauses per T-unit, type-token ratio, percentage of error-free clauses, and percentage of correct use of particles.

Table 31 displays the results of speech rate. A comparison between the NPG2 and PG2 groups revealed that the speech rate for PG2 was higher than that of the NPG2 group (NPG2 = 158.13 moras per minute, PG2 = 168.74 moras per minute). Regarding the results of the NOLP and the OLP groups, it was found that the NOLP group produced their stories at a higher speech rate than the OLP group (NOLP = 169.46 moras; OLP = 156.65 moras), which was expected since the participants in the OLP were encouraged to take time to think as they performed the task.

Table 31. Descriptive statistics for speech rate

<table>
<thead>
<tr>
<th>Pre-task planning</th>
<th>On-line planning</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>− Mean (SD)</td>
<td>176.64 (57.58)</td>
<td>158.13 (58.83)</td>
</tr>
<tr>
<td></td>
<td>+ Mean (SD)</td>
<td>162.27 (34.59)</td>
<td>168.74 (37.68)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>169.46 (47.55)</td>
<td>156.65 (51.9)</td>
</tr>
<tr>
<td></td>
<td>Adjusted mean</td>
<td>170.88</td>
<td>156.21</td>
</tr>
</tbody>
</table>

Regarding complexity, Tables 32 and 33 display the results for syntactic complexity and lexical complexity respectively. The results show that both the PG2 and
the NPG2 created stories with similar complexity. As for syntactic complexity, the NPG2 produced 1.90 clauses per T-unit, compared with 1.99 clauses per T-unit for the PG2. For lexical complexity, both groups produced .51 types per token.

For the NOLP and the OLP groups, the OLP group produced the stories with more clauses per T-unit than the NOLP group (NOLP = 1.86 clauses per T-unit, OLP = 1.95). In contrast, the task performance by the NOLP group showed higher TTR than that of the OLP group (NOLP = .49, OLP = .46).

Table 32. Descriptive statistics for the number of clauses per T-unit

<table>
<thead>
<tr>
<th></th>
<th>On-line planning</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td>Adjusted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SD)</td>
<td>(SD)</td>
<td>mean</td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>−</td>
<td>1.94</td>
<td>1.85</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>1.78</td>
<td>2.07</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>(.45)</td>
<td>(.56)</td>
<td>(.51)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>(.38)</td>
<td>(.60)</td>
<td>(.51)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.86 (SD)</td>
<td>1.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(.42)</td>
<td>(.58)</td>
<td></td>
</tr>
<tr>
<td>Adjusted mean</td>
<td>1.87</td>
<td>1.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 33. Descriptive statistics for type-token ratio

<table>
<thead>
<tr>
<th>Pre-task planning</th>
<th>On-line planning</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>− Mean (SD)</td>
<td>.53 (.08)</td>
<td>.51 (.09)</td>
</tr>
<tr>
<td></td>
<td>+ Mean (SD)</td>
<td>.46 (.08)</td>
<td>.45 (.08)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>.49 (.09)</td>
<td>.49 (.09)</td>
</tr>
<tr>
<td></td>
<td>Adjusted mean</td>
<td>.49</td>
<td>.46</td>
</tr>
</tbody>
</table>

Turning to accuracy, Table 34 displays the results of global accuracy and Table 35 displays the results of specific accuracy. It was found that the PG2 obtained somewhat higher accuracy rate for both measures, although the difference was less than 1% between the two groups (33.27% for error-free clauses and 88.53% for correct particles for the PG2 and 32.51% for error-free clauses and 87.97% for correct particles for the NPG2). Regarding the NOLP and the OLP groups, the OLP group produced narratives that were slightly more accurate than those of the NOLP group in terms of the percentage of error-free clauses (NOLP = 33.74%, OLP = 34.89%). The accuracy rate for correct particles is almost the same for the two groups (NOLP = 87.42%, OLP = 87.61%).
Table 34. Descriptive statistics for percentage of error-free clauses

<table>
<thead>
<tr>
<th></th>
<th>On-line planning</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>Mean (SD)</td>
<td>34.26 (13.79)</td>
<td>30.75 (18.28)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>33.21 (17.80)</td>
<td>39.61 (14.75)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td></td>
<td>33.74 (15.76)</td>
<td>34.89 (17.13)</td>
</tr>
<tr>
<td>Adjusted mean</td>
<td></td>
<td>33.96</td>
<td>34.51</td>
</tr>
</tbody>
</table>

Table 35. Descriptive statistics for percentage of correct use of particles

<table>
<thead>
<tr>
<th></th>
<th>On-line planning</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-task planning</td>
<td>Mean (SD)</td>
<td>88.72 (6.64)</td>
<td>87.22 (8.52)</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>86.12 (9.73)</td>
<td>88.04 (7.23)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td></td>
<td>87.42 (8.35)</td>
<td>87.61 (7.87)</td>
</tr>
<tr>
<td>Adjusted mean</td>
<td></td>
<td>87.47</td>
<td>87.50</td>
</tr>
</tbody>
</table>

These data on five measures for L2 oral production were subjected to a two-way MANCOVA in which pre-task planning and on-line planning were assigned as independent variables, and SPOT scores as covariates. The alpha level was set at .05.
Assumptions for a MANCOVA were evaluated, and it was found that they were not violated. The results of Wilks’ lambda test revealed a significant main effect for pre-task planning \((F = 2.76, p < .05)\), but not on-line planning \((F = 1.74, p = .135)\). The interaction effect between pre-task and on-line planning was not significant \((F = 1.97, p = .091)\). Thus, only the significant results that were found for pre-task planning in L2 learners’ oral production are considered in this study. The results of Wilks’ lambda are displayed in Table 36.

Table 36. Wilks’ lambda for a two-way MANCOVA for pre-task planning and on-line planning effects on independent variables

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>.859</td>
<td>2.76</td>
<td>5</td>
<td>84</td>
<td>.023*</td>
</tr>
<tr>
<td>On-line planning</td>
<td>.906</td>
<td>1.74</td>
<td>5</td>
<td>84</td>
<td>.135</td>
</tr>
<tr>
<td>Pre-task planning × On-line</td>
<td>.895</td>
<td>1.97</td>
<td>5</td>
<td>84</td>
<td>.091</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

In the following sections, the results of task planning effects on fluency, complexity, and accuracy in L2 oral production are presented for research questions 2.1, 2.2, and 2.3, followed by correlations among dependent variables for research question 2.4.
2. Are there any differences in the way pre-task and on-line planning affect the L2 learners’ oral production in a narrative task, controlling for pre-existing Japanese proficiency differences?

2.1 How do the different types of planning affect the fluency of L2 learners’ oral production in a narrative task?

2.2 How do the different types of planning affect the complexity of L2 learners’ oral production in a narrative task?

2.3 How do the different types of planning affect the accuracy of L2 learners’ oral production in a narrative task?

2.4 Are there any trade-off effects among the fluency, complexity, and accuracy of L2 learners’ oral production in a narrative task?

4.3.2.1 Fluency

Research question 2.1 asks how the different types of planning affect the fluency of L2 learners’ oral production in a narrative task. The results for this question are shown in Table 37. The covariate was significant for speech rate.

Table 37. Follow-up ANCOVA for task planning conditions on speech rate

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>2015.38</td>
<td>1.01</td>
<td>.319</td>
</tr>
<tr>
<td>On-line planning</td>
<td>4954.71</td>
<td>2.47</td>
<td>.120</td>
</tr>
<tr>
<td>Pre-task planning × On-line planning</td>
<td>12201.39</td>
<td>6.09</td>
<td>.016*</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
Concerning the effects of planning on fluency, no significant difference was found between the NPG2 and the PG2 \( (F = 1.01, p = .319) \). There is no significant difference in fluency between the NOLP and the OLP groups, either \( (F = 2.47, p = .120) \). The interaction between pre-task and on-line planning reached a significant level \( (F = 6.09, p = .016) \). However, as shown earlier, the Wilks’ lambda revealed no significant interaction effect, and therefore, this result is not considered.

### 4.3.2.2 Complexity

Research question 2.2 is concerned with the effects of different types of planning on the complexity of L2 learners’ oral production. In this section, the results of syntactic complexity are first reported, and then the results of lexical complexity.

#### 4.3.2.2.1 Syntactic complexity

With regard to syntactic complexity in L2 oral production, the results are shown in Table 38. The covariate was significant for this measure.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>.00</td>
<td>.01</td>
<td>.927</td>
</tr>
<tr>
<td>On-line planning</td>
<td>.08</td>
<td>.37</td>
<td>.542</td>
</tr>
<tr>
<td>Pre-task planning ×</td>
<td>.62</td>
<td>2.94</td>
<td>.090</td>
</tr>
<tr>
<td>On-line planning</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\* *p < .05, ** *p < .01
The analysis found no significant difference in syntactic complexity between the NPG2 and the PG2 ($F = .01, p = .927$) or between the NOLP and the OLP groups ($F = .037, p = .542$). Interaction between pre-task planning and on-line planning was not significant, either ($F = 2.94, p = .090$).

4.3.2.2.2 Lexical complexity

Table 39 displays the results of how different types of planning affected lexical complexity in L2 oral production. The covariate was not significant for lexical complexity.

Table 39. Follow-up ANCOVA for task planning conditions on TTR

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>.09</td>
<td>13.15</td>
<td>.000**</td>
</tr>
<tr>
<td>On-line planning</td>
<td>.03</td>
<td>3.69</td>
<td>.058</td>
</tr>
<tr>
<td>Pre-task planning × On-line planning</td>
<td>.00</td>
<td>.01</td>
<td>.934</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

The NPG2 outperformed the PG2, and Table 35 shows a statistically significant difference between the two groups ($F = 13.15, p = .000$). In other words, the participants in the NPG2 used a greater variety of vocabulary to tell their stories than did those in the PG2. This result contradicts those in past studies that revealed pre-task planning positively affected the lexical variety of L2 oral production (Gilabert, 2007; Kawauchi, 2005a, 2005b; Tajima, 2003). As for the NOLP and OLP groups, lexical complexity did
not reach significance although it is close to a significant level \((F = 3.69, p = .058)\). The interaction between pre-task and on-line planning did not attain significance either \((F = .01, p = .934)\).

**4.3.2.3 Accuracy**

Research question 2.3 inquires how the accuracy of L2 learners’ oral production is affected by planning. Table 40 shows the results of the analysis for global accuracy (percentage of error-free clauses), and Table 41 displays the results for specific accuracy (percentage of correct use of particles). The covariate was significant for both global and specific measures.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>236.27</td>
<td>1.11</td>
<td>.295</td>
</tr>
<tr>
<td>On-line planning</td>
<td>0.86</td>
<td>.00</td>
<td>.949</td>
</tr>
<tr>
<td>Pre-task planning × On-line planning</td>
<td>362.71</td>
<td>1.71</td>
<td>.195</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
Table 41. Follow-up ANCOVA for task planning conditions on the percentage of correct use of particles

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>24.21</td>
<td>.37</td>
<td>.544</td>
</tr>
<tr>
<td>On-line planning</td>
<td>0.02</td>
<td>.00</td>
<td>.986</td>
</tr>
<tr>
<td>Pre-task planning × On-line planning</td>
<td>53.36</td>
<td>.82</td>
<td>.368</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Table 40 shows that there is no significant difference between the NPG2 and PG2 ($F = 1.11, p = .295$) or between NOLP and OLP groups ($F = .00, p = .949$). The interaction was not significant either ($F = 1.71, p = .195$). Regarding specific accuracy, Table 41 shows that no significant difference was found across planning conditions ($F = .37, p = .544$ for pre-task planning and $F = .00, p = .986$ for on-line planning) or for the interaction between them ($F = .82, p = .368$).

4.3.2.4 Trade-off effects

For research question 2.4 which addresses trade-off effects among the fluency, complexity, and accuracy of L2 oral performance, the correlations were examined on five dependent variables for task performance across planning conditions (speech rate, number of clauses per T-unit, type-token ratio, percentage of error-free clauses, percentage of correct use of particles) for the participants in G1, G3, G5, and G6, and then within each of the four groups. Since the focus of the analyses is the trade-off effects, which means that as one condition improves, another declines, only negative correlations in the results are discussed in this section.
Table 42 displays the overall correlations for the five dependent variables. The results show that lexical complexity was negatively correlated with a specific accuracy measure at a low level ($r = -0.211$). Correlations among other variables were either negative at a non-significant level or were positive.

Table 42. Correlations among dependent variables for G1, G3, G5, and G6 (n = 93)

<table>
<thead>
<tr>
<th></th>
<th>Speech rate</th>
<th>Clauses/ T-unit</th>
<th>TTR</th>
<th>Error-free</th>
<th>Corr. Particles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses/T-unit</td>
<td>0.249*</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>-0.199</td>
<td>0.010</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>0.237*</td>
<td>0.419**</td>
<td>-0.211*</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Corr. Particles</td>
<td>-0.009</td>
<td>0.175</td>
<td>-0.05</td>
<td>0.373**</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Turning to correlations among dependent variables within each group, for G1, which performed the task without pre-task or on-line planning, no negative correlation at a significant level was found among the five dependent variables at a significant level. The results suggest that there may be no trade-off effect among fluency, complexity, and accuracy when L2 learners do not have opportunity to plan for a task. Table 43 displays the results for G1.
Table 43. Correlations among dependent variables for G1 (n = 24)

<table>
<thead>
<tr>
<th></th>
<th>Speech rate</th>
<th>Clauses/T-unit</th>
<th>TTR</th>
<th>Error-free</th>
<th>Corr. Particles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses/T-unit</td>
<td>.250</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>-.142</td>
<td>-.039</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>.167</td>
<td>.342</td>
<td>-.057</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Corr. Particles</td>
<td>-.080</td>
<td>.214</td>
<td>-.068</td>
<td>.244</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

G3 planned prior to the task performance, and the results for G3 are displayed in Table 44.

Table 44. Correlations among dependent variables for G3 (n = 24)

<table>
<thead>
<tr>
<th></th>
<th>Speech rate</th>
<th>Clauses/T-unit</th>
<th>TTR</th>
<th>Error-free</th>
<th>Corr. Particles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses/T-unit</td>
<td>.078</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>-.097</td>
<td>-.214</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>.146</td>
<td>.305</td>
<td>-.527**</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Corr. Particles</td>
<td>.015</td>
<td>.003</td>
<td>-.285</td>
<td>.440*</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
For G3, a moderate negative correlation was found between TTR and percentage of error-free clauses, which suggests a trade-off between lexical complexity and overall accuracy for L2 learners’ oral production ($r = -0.527^{**}$). A similar trend was observed for G5 as seen in Table 45, although the correlation did not reach a significant level ($p = .087$). It appears that a trade-off effect may occur between lexical complexity and accuracy when L2 learners have an opportunity to plan either before or during task performance.

In addition to the percentage of error-free clauses, TTR for G5 was negatively correlated with speech rate at a moderate level ($r = -0.442^*$). For G1 and G3, no correlation was found between the two variables and, therefore, there may be a trade-off between lexical complexity and fluency as well as between lexical complexity and overall accuracy when L2 learners plan during the task.

Table 45. Correlations among dependent variables for G5 (n = 24)

<table>
<thead>
<tr>
<th></th>
<th>Speech rate</th>
<th>Clauses/ T-unit</th>
<th>TTR</th>
<th>Error-free</th>
<th>Corr. Particles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses/T-unit</td>
<td>.349</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>-0.442*</td>
<td>.088</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>.262</td>
<td>.415*</td>
<td>-0.357</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Corr. Particles</td>
<td>-0.054</td>
<td>.220</td>
<td>-0.045</td>
<td>.260</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*$p < .05$, **$p < .01$
Finally, correlations among five dependent variables for G6 were examined, and the results are displayed in Table 46.

<table>
<thead>
<tr>
<th></th>
<th>Speech rate</th>
<th>Clauses/ T-unit</th>
<th>TTR</th>
<th>Error-free</th>
<th>Corr. Particles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses/T-unit</td>
<td>.128</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>-.055</td>
<td>.249</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>.259</td>
<td>.551**</td>
<td>.484*</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Corr. Particles</td>
<td>.017</td>
<td>.220</td>
<td>.211</td>
<td>.571**</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

The results revealed no negative correlation between any pairs of variables, which suggests that there is no trade-off among fluency, complexity, and accuracy when L2 learners plan both before and during the task. It is interesting that the TTR was positively correlated with the percentage of error-free clauses for G6, whereas negative correlation between these two variables was observed for G3 and G5. It appears that lexical complexity and accuracy negatively affect each other only when L2 learners plan either before or during the task performance, but not when they plan both before and during the task.

4.4 Summary of quantitative results

In this chapter, results of quantitative analyses of participants’ speech data were reported in response to two research questions regarding the effects of time pressure and
planning opportunities on L2 oral production. To analyze the data, a factor analysis was first performed to determine the most appropriate measures for fluency, complexity, and accuracy of narrative stories for this study, and five measures were selected. Using these measures as dependent variables, the effects of time pressure on L2 oral production were analyzed to answer research question 1. The analysis revealed that participants who performed the task under time pressure spoke significantly faster than those who had no time pressure during the task. Neither complexity nor accuracy was affected by time pressure.

To address the second research question, the effects of different types of planning on narrative performance were examined. The results revealed that the participants who had no pre-task planning opportunity produced narrative stories with a greater variety of vocabulary than those who planned before the task. There was no significant difference in fluency, syntactic complexity, or accuracy across planning types. Finally, the correlations among the five dependent variables were examined for trade-off effects. Overall, TTR was negatively correlated with the percentage of error-free clauses to a minor degree. Within-group comparisons also show a negative correlation between TTR and percentage of error-free clauses for G3 at a significant level and for G5 at a near-significant level, suggesting a trade-off effect between lexical complexity and accuracy when L2 learners plan either before or during the task. TTR was negatively correlated with speech rate for G5 as well.

In the next chapter, the analyses of retrospective interview data are presented to address research question 3, which asked about L2 learners’ strategy use during task planning and performance.
CHAPTER 5: RESULTS OF INTERVIEW ANALYSES

5.1 Introduction

The third research question asks what L2 learners do during planning. To answer the question, data were collected through interviews that consisted of a set of predetermined questions and stimulated recall (see Appendix D). The strategies that the participants used were classified into three categories of strategies: metacognitive, cognitive, and social/affective strategies (see section 3.6.2.1).

In this portion of the study, a total of 24 Japanese learners in Schools A and B, four learners each from six experimental groups of G1 through G6, participated. The interviews for G2 and G4, who performed the task with a time limit, included some questions regarding time pressure, which offers additional information on research question 1 on the effects of time pressure on task performance.

In this chapter, first, the perspectives of the participants in G2 and G4 on time pressure are presented. Then, the strategies that were used by the participants in G1, G3, G5 and G6 during task planning and task performance are reported in the following order: (a) selection of strategies by planning conditions; (b) analyses of strategy types; and (c) depth of language analyses with or without planning opportunities.

5.2 Effects of time pressure

In chapter 4, the effects of time pressure on participants’ task performance were examined, and it was found that the Time Pressure group that comprised G2 and G4 spoke significantly faster than the No Time Pressure group (G1 and G3), but no differences were found between the two groups in the complexity or accuracy of their speeches. In this section, the interview data from eight participants in the Time Pressure group (4 participants each from G2 and G4) were reviewed in terms of how these participants perceived the time pressure in order to further investigate the effects of time pressure on L2 oral production (research question 1). As explained in chapter 3, the
participants in G2 performed the task after 30 seconds from the time they received the pictures for the task, whereas those in G4 had an opportunity to plan for the task for 10 minutes before the task.

The interview questions that are considered for this analysis are as follows: (a) how participants felt about the time limit, and (b) whether the participants reported that their stories would have been different if there had not been a time limit. When the participants said that their stories would have been different in response to the second question, they were further asked in what way they thought that their story was affected by the time limit. The results are presented in Table 47 by combining the responses from the participants in both G2 and G4.

Table 47. Participants’ perceptions of time pressure (n = 8)

<table>
<thead>
<tr>
<th>Concerned</th>
<th>4</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Content</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Grammar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To the question about participants’ perceptions of the impact of time pressure, of the eight participants in the Time Pressure group, half of them mentioned some kind of effects from the time pressure, whereas the other half stated that they were not concerned about the time limit while performing the task. Of the four participants who were
concerned about the time pressure, three indicated that they felt hurried due to the time pressure. The following is an excerpt from one of these participants. The first number represents the group to which this participant belonged.

(2) G4–4: I felt it was kind of short, so I felt, I felt a little bit hurried.

The other participant expressed that the time pressure made her nervous.

In response to the question that asked participants in the Time Pressure group whether their story would have been different if there had not been a time limit, all but one of the participants answered that they altered their stories due to the time limit. For these seven participants who felt that their stories were affected by the time pressure, it was further asked how their stories would have been different without the time limit. Four participants mentioned the content of their stories, whereas three participants talked about grammar. Of the four participants who answered that the content of their stories were affected by the time limit, one participant said the following:

(3) G4–3: Probably you can go into more detail, just have to figure out the more important points and things like that.

---

1 In all the excerpts from the interview, Japanese words are italicized, and their translations follow immediately in square brackets. The participants are identified with their group numbers plus individual numbers that were given at the time of data collection. Comments by the interviewer appear in parentheses that start with R, which indicates the researcher.
Similar to G4–3, another participant in the Time Pressure group mentioned that she would have been able to add more content to her story without the time pressure. Another participant thought his story would have been more organized, and still another participant stated that he would have been able to think about the story more thoroughly.

Another group of participants in the Time Pressure group stated that the grammar in their stories was altered due to the time pressure. The following example is a case where a participant described how the grammar in his story would have been different without the time limit.

(4) G2–2: I might have been able to use better tenses, forms in the verbs, and connect the sentences.

In addition to example (4), one participant stated that she might have been able to think of more grammar points without the time pressure, while the other participant mentioned that she would have been able to use more accurate grammar in her story.

It should be noted that the participants in the Time Pressure group responded similarly in terms of whether or not their stories would have been different without the time pressure regardless of the provision of pre-task planning. However, some differences were observed between G2 and G4 in their perceptions of how their stories would have been different. Of the four participants who responded that the content of their stories was affected by the time pressure, three were participants in G4, and only one participant was in G2. In contrast, all three participants who answered that the time pressure influenced the grammar in their stories were in G2. Since there were only four participants each in G2 and G4 whose interviews were analyzed for this study, individual differences should be assumed. However, it is worth pointing out that the participants who were able to spend 10 minutes to prepare for the task felt that the content of their
stories would have been improved without the time pressure, whereas those who had no opportunity to plan before the task thought that they would have been able to improve the grammar in their stories if there had been no time pressure. The participants in G4 thought that the content of their stories suffered because they had to finish the stories within one and a half minutes and, consequently, their stories were simplified, not because they did not have enough time to think about their stories. In the case of the participants in G2, they had to start telling the story almost immediately after they received the pictures for the task and, therefore, they had no time to think about the task, either the grammar or the content, prior to their task performance. Due to the time limit, they did not have much time to think about the task during the task performance, either. Yet, the participants in G2 mentioned only the grammar, stating that their grammar would have been better without the time limit. These results suggest that the participants might have been more inclined to be concerned about grammar over content of the task when they had no opportunity to plan before and during the task.

5.3 Strategy use

5.3.1 Introduction

In response to research question 3 regarding L2 learners’ strategy use during planning and task performance, the average number of strategy types that participants in G3, G4, and G6 (the Pre-Task Planning group) used during the pre-task planning, as well as G1 through G6 used during the task, was computed for each of the three strategy categories: metacognitive, cognitive, and social/affective strategies. The number of strategy types that were used for on-line planning was calculated separately for the On-line Planning group (G5 and G6) and the No On-line Planning group (G1, G2, G3, and G4) to examine the differences in strategy use when the participants were encouraged or not encouraged to plan while performing the task. Since each planning group consists of the experimental groups that received different planning conditions, the analysis includes
comparisons within individual planning groups if any differences were observed among the sub-groups.

Strategy types for this study are presented in Tables 8–10 in chapter 3. The data were coded for three categories of learner strategies based on a modified version of Ortega’s (2005) classifications. Originally, self-monitoring strategy was divided in several subcategories: Production monitoring, Auditory monitoring, Visual monitoring, Cross-language monitoring, Style and register monitoring, Double-check monitoring, and Strategy monitoring. However, after examining the interview data in this study, it was found that participants did not use various monitoring strategies, and therefore, all the monitoring strategies were combined as Self-monitoring. A similar trend was observed for evaluation strategies. Five types of evaluation strategy were listed in Ortega (2005) (Performance evaluation, Repertoire evaluation, Ability evaluation, Strategy evaluation, and Prognostic evaluation), but not all the evaluation strategies were used by the participants in this study, and they were combined as Self-evaluation. Thus, participants’ use of metacognitive strategies were analyzed for five types: 1.1. Advance planning, 1.2. Organizational planning, 2. Problem identification, 3. Self-monitoring, 4. Self-evaluation.

5.3.1.1. Strategy use during pre-task planning

In this section, the strategy use during pre-task planning is examined. The participants in the Pre-Task Planning group are those in G3, G4, and G6; that is, the learners who had the opportunity to plan before the task performance. The results of strategy use by individual groups with the pre-task planning opportunity as well as the Pre-Task Planning group as a whole are displayed in Table 48. The ratio of the number of strategy types that were used by the participants for each of the three strategy categories (5 types for metacognitive strategies, 14 types for cognitive strategies, and 3 types for social/affective strategies) is shown in the parentheses in the table.
Table 48. Strategy use by the Pre-Task Planning group before the task

<table>
<thead>
<tr>
<th></th>
<th>MC (5 types)</th>
<th>C (14 types)</th>
<th>S/A (3 types)</th>
<th>Total (32 types)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G3</td>
<td>4 (80%)</td>
<td>3 (21.43%)</td>
<td>0.25 (8.33%)</td>
<td>7.25</td>
</tr>
<tr>
<td>G4</td>
<td>3.75 (75.00%)</td>
<td>3.75 (26.79%)</td>
<td>0 (0%)</td>
<td>7.5</td>
</tr>
<tr>
<td>G6</td>
<td>3.75 (75.00%)</td>
<td>3.25 (23.21%)</td>
<td>0.25 (8.33%)</td>
<td>7.25</td>
</tr>
<tr>
<td>Mean</td>
<td>3.83 (76.60%)</td>
<td>3.33 (23.79%)</td>
<td>0.17 (5.67%)</td>
<td>7.33</td>
</tr>
</tbody>
</table>

The results show that generally the participants in the Pre-Task Planning group used a similar variety of strategies for metacognitive and cognitive strategies: 3.83 types for metacognitive strategies and 3.33 types for cognitive strategies. The social/affective strategies were hardly used by this group (0.17 types).

The participants in G3, G4, and G6 were provided with different instructions for task performance. G4 had a time limit to perform the task whereas G6 was encouraged to take time to think while telling the stories, and G3 received no instruction for how to perform the task. Despite these differences in task performance conditions, the participants in these groups showed similar patterns of strategy use. All of the groups used a similar variety of metacognitive and cognitive strategies and used almost no social/affective strategies. Also, the range of the number of strategy types that the participants used for each of metacognitive, cognitive, and social/affective strategies is relatively small across groups (3.75–4 types for metacognitive strategies, 3–3.75 types for cognitive strategies, and 0–0.25 for social/affective strategies). Therefore, it appears
that the task performance conditions did not affect how the participants planned before the task.

5.3.1.2. Strategy use during task performance

This section focuses on the participants’ strategy use during the task. To identify any differences in the selections of strategies depending on the opportunity of conscious on-line planning, the interview data were analyzed individually for the No On-line Planning group (G1–G4) and the On-line Planning group (G5 and G6). For the task performance, the participants in the On-line Planning group were encouraged to think as they performed the task whereas those in G1–G4 were not. In addition, the On-line Planning group had an opportunity to practice planning on-line during the familiarization task. Among the groups that were involved in this comparison, G1, G2, and G5 had 30 seconds to look through the pictures before telling the story. The strategies that the participants in these groups used during this period are included in the analyses for this study.

Tables 49 and 50 illustrate the results for the No On-line Planning group and for the On-line Planning group. The participants in the No On-line Planning group used 5.44 types in total. The results showed that they employed the greatest variety in metacognitive strategies (3.63 types), followed by cognitive strategies (1.69 types). The social/affective strategies were hardly used (0.13 types).

As with the On-line Planning group, the results for this group generally echoed those for the No On-line Planning group; that is, they had the most variety in metacognitive strategies (4 types), followed by cognitive strategies (1.88 types). The least variety of strategies was used for social/affective strategies among three categories (0.75 types). However, a few differences were found between the On-line Planning group and No On-line Planning group. First, the On-line Planning group used 6.63 types of strategies in total whereas the No On-line Planning group used 5.44 types. The On-line
Planning group, therefore, used more variety of strategies than the No On-line Planning group by approximately 1.2 types. Also, the participants in the On-line Planning group used 0.75 types of social/affective strategies. The number may seem small, but it is more than five times greater than the participants in the No On-line Planning group did (0.13 types). Thus, on-line planning seemed to encourage the use of social/affective strategies.

Table 49. Strategy use by the No On-line Planning group during task performance

<table>
<thead>
<tr>
<th>Group</th>
<th>MC (15 types)</th>
<th>C (14 types)</th>
<th>S/A (3 types)</th>
<th>Total (32 types)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>4 (80%)</td>
<td>1.75 (12.5%)</td>
<td>0 (0%)</td>
<td>5.75</td>
</tr>
<tr>
<td>G2</td>
<td>3.75 (75%)</td>
<td>1.25 (8.93%)</td>
<td>0 (0%)</td>
<td>5</td>
</tr>
<tr>
<td>G3</td>
<td>3.75 (75%)</td>
<td>2 (14.29%)</td>
<td>0 (0%)</td>
<td>5.75</td>
</tr>
<tr>
<td>G4</td>
<td>3 (60.0%)</td>
<td>1.75 (12.5%)</td>
<td>0.5 (16.67%)</td>
<td>5.25</td>
</tr>
<tr>
<td>Mean</td>
<td>3.63 (72.60%)</td>
<td>1.69 (12.07%)</td>
<td>0.13 (4.33%)</td>
<td>5.45</td>
</tr>
</tbody>
</table>
Table 50. Strategy use by the On-line Planning group during task performance

<table>
<thead>
<tr>
<th></th>
<th>MC (15 types)</th>
<th>C (14 types)</th>
<th>S/A (3 types)</th>
<th>Total (32 types)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G5</td>
<td>4 (80%)</td>
<td>2.75 (19.64%)</td>
<td>0.5 (16.67%)</td>
<td>7.25</td>
</tr>
<tr>
<td>G6</td>
<td>4 (80%)</td>
<td>1 (7.14%)</td>
<td>1 (33.33%)</td>
<td>6</td>
</tr>
<tr>
<td>Mean</td>
<td>4 (80%)</td>
<td>1.88 (13.43%)</td>
<td>0.75 (25%)</td>
<td>6.63</td>
</tr>
</tbody>
</table>

The participants in the No On-line Planning group performed the task under different planning conditions as well as different task conditions depending on the experimental groups to which they were assigned. The participants in G1 and G2 started telling the story after 30 seconds from the time they received the pictures for the task, whereas those in G3 and G4 had 10 minutes to plan before the task performance. Another difference was the provision of a time limit. G2 and G4 had to finish the task within one and a half minutes, but G1 and G3 had no time limit for the task. However, the strategy use among these groups seemed to be generally identical regardless of the differences in planning and task conditions. All four groups used the greatest variety of metacognitive strategies, followed by cognitive strategies, and hardly used social/affective strategies. The number of types of strategies that each group used is similar to one another, ranging from 5 types to 5.75 types in total.

The On-line Planning group comprises G5 and G6 and there are some differences in the variety of strategies that these two groups used. The number of metacognitive strategies that G5 and G6 used is the same. Both groups used 4 types of metacognitive strategies. On the other hand, regarding the cognitive strategies and social/affective strategies, there were some differences between G5 and G6. G5 used more cognitive
strategies than G6 (2.75 types and 1 type respectively), whereas G6 used twice as many social/affective strategies (1 type and 0.5 types respectively). The participants in both groups were encouraged to think while telling the story, but G5 had to start the task almost immediately after receiving the pictures for the task, whereas G6 had 10 minutes to plan before the task. It is possible that the provision of the pre-task planning contributed to these differences. There are only four participants in each group, and therefore, participants’ individual differences may have affected the results, but these differences are worth pointing out.

5.3.1.3. Comparison between pre-task planning and on-line planning

Comparing the strategy use between pre-task planning and on-line planning, the results revealed that two groups with planning opportunities used more types of strategies than the group without such opportunity. The participants in the Pre-Task Planning group used the greatest variety of strategy types in total, followed by the On-line Planning group: 7.33 types for the Pre-Task Planning group and 6.63 types for the On-line Planning group. The No On-line Planning group used the least variety with 5.44 types. Although the Pre-Task Planning group used the most variety of strategies among these three groups, the difference between this group and the other two is relatively small considering that the participants in this group had 10 minutes to plan before the task performance.

Regarding three categories of strategies, overall the participants in all three groups showed the same pattern in spite of the differences in the planning conditions; that is, they used a greater number of strategy types for metacognitive strategies than cognitive strategies and used the least variety of social/affective strategies. Concerning metacognitive strategies, the On-line Planning group used the largest variety of the strategies, but overall the participants used a similar number of types of metacognitive
strategies across the planning conditions: 3.83 types for the Pre-Task Planning group (76.67%), 4 types for the On-line Planning group (80%), and 3.63 types for the No On-line Planning group (72.50%).

With regard to cognitive strategies, the difference among groups was larger than for the metacognitive strategies. The Pre-Task Planning group used the most variety of strategies with 3.33 types (23.79% of the cognitive strategies) while the On-line Planning group used 1.88 types (13.43%), and the No On-line Planning group used 1.69 types (12.07%), both of which are less than two-thirds of the variety for the Pre-Task Planning group and are less than half of the number of metacognitive strategies that the same groups used. These results show that the participants used more variety of cognitive strategies during the pre-task planning than during the on-line planning and that the on-line planning conditions did not affect the diversity of the cognitive strategies used while performing the task. Considering that the participants used a similar number of strategy types for metacognitive strategies across planning conditions, the use of cognitive strategies may be more inclined to change according to the timing of the planning and decrease during the task performance.

With regard to social/affective strategies, only three types of strategies were considered in this study; generally speaking, the participants used hardly any of these strategies. The On-line Planning group used the largest variety of social/affective strategies, but it was still less than one type on average (0.75 types). The Pre-Task Planning group used 0.17 types, whereas the No On-line Planning group used 0.13 types of the strategies in this category. Therefore, the results indicate that the participants may use social/affective strategies when they have an opportunity to plan on-line, but they use a limited variety of these strategies.
5.3.2 Selection of strategies by planning conditions

The previous section showed that despite the differences in planning conditions, participants generally used more types of metacognitive strategies than cognitive strategies, and they hardly used social/affective strategies. This section further analyzes the results of participants’ strategy use and compares whether there were any differences in the selection of strategies within three categories of strategies depending on the planning conditions. To examine the strategies that were used and were not used, the strategies that were used by half or more participants as well as those that were used less than a quarter of the participants in each planning group are analyzed. Just as in the previous section, the Pre-Task Planning group consists of the participants in G3, G4, and G6, who had an opportunity to plan for the task for 10 minutes before task performance. The On-line Planning group comprises the participants in G5 and G6 who were instructed to plan as they told their stories whereas the No On-line Planning group consisted of the participants in G1, G2, G3, and G4 who received no instruction for on-line planning. The results for on-line planning included the strategies that the participants in G1, G2, and G5 used during the 30-second period they had before the task.

5.3.2.1. Metacognitive strategies

In this section, participants’ use of metacognitive strategies is reported by planning condition. The strategies used during pre-task planning are shown in Table 51. The numbers in parentheses show the ratio of the number of participants who used the strategy in the group (12 participants in the Pre-Task Planning group, 8 participants in the On-line Planning group, and 16 participants in the No On-line Planning group).
Table 51. Participants’ use of metacognitive strategies by group

<table>
<thead>
<tr>
<th></th>
<th>Pre-task (n = 12)</th>
<th>On-line (n = 8)</th>
<th>No On-line (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Advance planning</td>
<td>12 (100%)</td>
<td>3 (37.5%)</td>
<td>7 (43.75%)</td>
</tr>
<tr>
<td>1.2. Organizational planning</td>
<td>12 (100%)</td>
<td>8 (100%)</td>
<td>12 (75%)</td>
</tr>
<tr>
<td>2. Problem identification</td>
<td>9 (75%)</td>
<td>8 (100%)</td>
<td>15 (93.75%)</td>
</tr>
<tr>
<td>3. Self-monitoring</td>
<td>7 (58.33%)</td>
<td>6 (75%)</td>
<td>15 (93.75%)</td>
</tr>
<tr>
<td>4. Self-evaluation</td>
<td>6 (50.00%)</td>
<td>7 (87.5%)</td>
<td>7 (43.75%)</td>
</tr>
</tbody>
</table>

The results show that all the participants in the Pre-Task Planning group used Advance planning and Organizational planning strategies. The majority of the participants in this group used Problem identification as well (9 of 12 participants, 75% of the participants in the group). The rest of the metacognitive strategies were used by less than half of the participants in the Pre-Task Planning group.

With respect to the strategies that were used during the task, all the participants in the On-line Planning group used Organizational planning and Problem identification. Self-monitoring and Self-evaluation were also used by many participants in the On-line Planning group: 6 participants (75%) for Self-monitoring and 7 participants (87.5%) for Self-evaluation.

Turning to the strategies used by the No On-line Planning group, the selection of strategies was somewhat similar to that of the On-line Planning group. Of 16 participants in the No On-line Planning group, 15 participants (93.75%) used Problem identification
and Self-monitoring. Organizational planning was also used by many of the participants in the No On-line Planning group (12 participants, 75%).

Comparing the three planning groups, the participants in these three groups used largely similar strategies. The majority of the participants in these groups used Organizational planning and Problem identification. A few differences in strategy selections were observed, however. One of these differences was the use of Advance planning. All of the participants in the Pre-Task Planning group used this strategy during the 10-minute planning time, whereas less than half of the participants in both the On-line and No On-line groups did during the task performance (3 participants, 37.5%, and 7 participants, 43.75%, respectively). It should be noted that the three participants in the On-line Planning group who used Advance planning were all in G5, and they used this strategy during the 30-second period before they started talking. All of the participants in G6 used Advance planning during pre-task planning, but not during the task. Similarly, the seven participants in the No On-line Planning group who used the Advance planning on-line were the participants in G1 and G2. All of the participants in G3 and G4 used it during pre-task planning, but again, not during the task. Just as the participants in G6 in the On-line Planning group who had 10 minutes before the task, all of the participants in G3 and G4 used this strategy during pre-task planning.

Self-monitoring was used by more than half of the participants in each group, although the strategy seems to be used by more participants during task performance regardless of the availability of on-line planning: 7 participants in the Pre-Task Planning group (58.33%), 6 participants in the On-line Planning group (75%), and 15 participants in the No On-line Planning group (93.75%). The results indicate that the participants more likely used this strategy when the task is actually carried out rather than when preparing for the task.

As mentioned earlier, this study originally attempted to classify participants’ strategy use in several types of monitoring strategies (Table 8), and it was found that the
The majority of the participants who used the Self-monitoring strategy in the On-line and No On-line Planning groups used Production monitoring: all 6 participants who used Self-monitoring strategy in the On-line Planning group (75%) and 14 of 15 participants who used Self-monitoring strategy in the No On-line Planning group (87.5%). In contrast, only two participants in the Pre-Task Planning group used Production monitoring. More participants in this group used Style and register monitoring and Double-check monitoring although the number of participant who used these strategies were less than half (3 participants, 25%, and 4 participants, 33.33%, respectively). Style and register monitoring was used by 7 participants in the No On-line Planning group as well, though it is also less than half of the participants in the group (43.75%). The participants in the On-line Planning group did not use any other types of monitoring strategies, except that one participant used Style and register monitoring. Considering these results, it appears that the participants monitored their own speech in a rather general manner during task performance, but not in more specific fashion, particularly when they were encouraged to plan on-line. During the pre-task planning, more participants might have used specific monitoring strategies, but not to the extent that shows any particular patterns.

In the same way with Self-monitoring, the data were analyzed for several types of Self-evaluation strategies at first, but they were combined as Self-evaluation because most of the types of evaluation strategies were not used by the participants in this study. Among three planning groups, the On-line planning group showed the highest percentage for the number of participants who used Self-evaluation (7 participants, 87.5%), and the majority of them used Performance evaluation (5 participants, 62.5%). Less than half of the participants in the No On-line Planning used Self-evaluation (7 participants, 43.75%), but similar to the participants in the On-line Planning, most of the participants in No On-line Planning group who used Self-evaluation engaged in Performance evaluation (6 of 7 participants). Thus, it seems that the participants evaluated their own speeches in a general manner as they described the story, particularly when they had enough time for
on-line planning. On the contrary, only one participant in the Pre-Task Planning group used Performance evaluation (8.33%), although a half of the participants in the group used some types of Self-evaluation. These participants used different types of evaluation strategies, and there was no specific type of evaluation strategies that was used by the majority of the participants who employed Self-evaluation.

5.3.2.2. Cognitive strategies

In respect of cognitive strategies, the results show that the participants generally used less variety of cognitive strategies than they did for metacognitive strategies regardless of the planning conditions (Table 52). Regarding the Pre-Task Planning group, it was found that Rehearsal was the most commonly used strategy, which was used by nine participants (75%), followed by Outlining/summarizing, which was used by six participants (50%). All the other cognitive strategies were used by fewer than half of the participants.

With regard to the strategy use for cognitive strategies during task performance, it was generally similar between the On-line Planning and No On-line Planning groups, and there was only one strategy that was used by half of the participants in both groups, which was Approximating. This strategy was used by four participants in the On-line Planning (50%) and eight participants in the No On-line planning group (50%). None of the other cognitive strategies were used by the On-line Planning group nor by the No On-line Planning group during the task.
Table 52. Participants’ use of cognitive strategies by group

<table>
<thead>
<tr>
<th></th>
<th>Pre-task (n = 12)</th>
<th>On-line (n = 8)</th>
<th>No On-line (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Writing for retrieval</td>
<td>2 (16.67%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>1.2. Elaboration</td>
<td>3 (25%)</td>
<td>2 (25%)</td>
<td>1 (6.25%)</td>
</tr>
<tr>
<td>2.1 Writing for later recall</td>
<td>4 (33.33%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>2.2. Rehearsing</td>
<td>9 (75%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>3.1. Highlight and postpone</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>3.2. Make up</td>
<td>2 (16.67%)</td>
<td>3 (37.5%)</td>
<td>6 (37.5%)</td>
</tr>
<tr>
<td>4. Avoidance</td>
<td>1 (8.33%)</td>
<td>1 (12.5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>5.1. Approximating</td>
<td>4 (33.33%)</td>
<td>4 (50%)</td>
<td>8 (50%)</td>
</tr>
<tr>
<td>5.2. Circumlocution and synonyms</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (6.25%)</td>
</tr>
<tr>
<td>5.3. Switching to the mother tongue</td>
<td>3 (25%)</td>
<td>1 (12.5%)</td>
<td>6 (37.5%)</td>
</tr>
<tr>
<td>5.4 Other lexical compensation</td>
<td>2 (16.67%)</td>
<td>3 (37.5%)</td>
<td>3 (18.75%)</td>
</tr>
<tr>
<td>6. Translating</td>
<td>3 (25%)</td>
<td>1 (12.5%)</td>
<td>2 (12.5%)</td>
</tr>
<tr>
<td>7. Across-language analysis</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>8. Outlining/summarizing</td>
<td>6 (50%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>
Unlike the selections of metacognitive strategies, the cognitive strategies that were utilized during pre-task planning and on-line planning are different, and none of the cognitive strategies were used both during pre-task and on-line planning. The Pre-Task Planning group used Rehearsing and Outlining/summarizing to prepare for the task, but neither the On-line nor No On-line Planning group used these strategies during the task. It is highly likely that both of these strategies were prompted by the pre-task planning conditions. The participants in the Pre-Task Planning group made use of planning time by outlining and summarizing their stories. In addition, they had time to practice telling the stories before starting to perform the task. In contrast, the participants in the On-line and No On-line Planning group used Approximating, which was not used by most of the participants in the Pre-Task Planning group (4 participants, 33.33%). The results suggest that the decisions of altering the stories seemed to happen more often during the task performance. Just as with metacognitive strategies, many of the cognitive strategies were used by less than 25% of the participants in each planning group. There were 14 types of cognitive strategies, and the participants in the Pre-Task Planning group used 10 of them, whereas those in both the On-line Planning and No On-line Planning groups used 11 of them.

5.3.2.3. Social/affective strategies

As to social/affective strategies, Table 53 displays the results. It was found that only Encouraging oneself was used by as many as half of the participants in the On-line Planning group. Although there were other strategies in the metacognitive and cognitive strategies that were used by more participants, the use of this strategy is worth pointing out, considering that most of the participants in other planning condition groups did not use any of the social/affective strategies. It appears that the participants had a need to attend to their emotional sides during the task performance and that they were able to
respond to the need only when they were allowed to take time during the task performance.

Table 53. Participants’ use of social/affective strategies by group

<table>
<thead>
<tr>
<th></th>
<th>Pre-task (n = 12)</th>
<th>On-line (n = 8)</th>
<th>No On-line (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lowering anxiety</td>
<td>0 (0%)</td>
<td>1 (12.5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>2. Encouraging oneself</td>
<td>0 (0%)</td>
<td>4 (50%)</td>
<td>1 (6.25%)</td>
</tr>
<tr>
<td>3. Empathizing with the listener</td>
<td>2 (16.67%)</td>
<td>1 (12.5%)</td>
<td>1 (6.25%)</td>
</tr>
</tbody>
</table>

5.3.3 Depth of language analyses

In the previous sections, the types of strategies that the participants in this study used before and during the task were observed, and it was found that they generally used similar strategies regardless of their planning condition, though there were some differences. However, despite the similarities regarding the types of strategies that were used by the participants, there seemed to be some differences in terms of how deeply the participants with and without planning opportunities were engaged in language processing to retrieve their L2 knowledge. It is particularly noticeable when the participants encountered some difficulties in performing the task. For example, when participant G1-3, who had no opportunity to plan either before or during the task, did not know a word for “to turn on” in Japanese, she decided to guess, as shown in example (5) below.
(5) G1-3 (on-line): I couldn’t remember the word for “to turn
on,” so I was like, “Uh, I can’t remember,” so I just guessed,
and I stumbled through it.

Some participants decided to use English to compensate for unknown Japanese
word(s). The following is the example for such a case. Participant G1-1 did not know the
Japanese word for a “plug,” and she decided to say puragu, making an English word
“plug” sound like a Japanese word.

(6) G1-1 (on-line): I couldn’t think of a way to say … “put the
power cord into the plug” without using katakana. So, I ended
up using puragu [plug].

These types of compensation strategies were used by some participants with
planning opportunities as well, whether pre-task, on-line, or both pre-task and on-line, but
the participants in these groups also used other strategies that required deeper thinking to
try to remember the words that they needed to tell the story. Consider the following
example of G6-2, who tried to think of the word he wanted by using his knowledge of
other terms.

(7) G6-2 (pre-task) I tried to figure out how I would say that
the computer wouldn’t turn on, but I wasn’t sure ’cause I knew
how to say the light wouldn’t turn on. I wasn’t sure if you’d
use the same word for a computer or not.
Likewise, one participant in G5 tried to trace back his memory from his personal experience at a similar situation when telling the story about the second picture, which shows that the main character sat down in front of the computer.

(8) G5-1 (on-line): I’m really just trying to think back, trying to bring myself to a time where this might have been a similar situation. I was trying to think of…the words I might have used or somebody might have…said in Japanese.

(R: In that particular situation?)

Right. So I was thinking, often times somebody would ask,

\textit{Kono seki wa aiteimasu ka? [Is this seat available?]}

One of the other types of examples that often appeared in the interviews for the participants with planning opportunities was a comparison of two possible words. The participants tried to identify which word would be a better fit for the story. To describe that the main character told a person at a help desk about a computer problem, G3-3 reported that he was thinking on-line if he should say “to tell” or “to explain.”
(9) G3-3 (on-line): I was kind of wondering whether I should use oshieru [to tell] as opposed to setsu [setsumei, to explain]… I wasn’t… oshieru [to tell] overall was probably a better choice.

(R: You mean, to tell the teacher about the problem?)

Yeah, to inform, instead of just explain.

There were some cases in which the participants sought words or expressions even when they knew at least one way to describe the pictures. In the following example, G5-2 knew that she could use tonari no hito [the person next to] to describe a student who was sitting next to the main character in the picture. However, she further thought about possible ways to explain the person.

(10) G5-2 (on-line): This guy who is sitting at the computer next to them through the whole thing, it just seemed easiest to say tonari no hito [the person next to], but…I was thinking about how am I going to describe this person.

The above examples are some of the typical comments from the participants with any types of planning, but the same trend was not found in the interview data in G1.

5.4 Summary of the results of interview analyses

This chapter first examined the interview data to investigate how participants with time pressure (G2 and G4) felt during the task performance. It was found that the
majority of these participants felt that their stories were affected by time pressure. However, the perceptions on how their stories were affected were different between G2 and G4. The participants in G2 thought that their grammar was affected whereas those in G4 felt that the contents of their stories were affected.

The remaining part of this chapter analyzed the participants’ task-planning strategies to answer research question 3, which asks what L2 learners do during planning. Reviewing the results of the strategy use by each group revealed that the participants used similar strategies regardless of their planning conditions. All the groups used metacognitive strategies the most both before and during the task, followed by cognitive strategies. However, there were some differences in the selection of cognitive strategies between pre-task planning and on-line planning phases.

Although the participants across planning conditions used largely similar strategies, there were some differences in the depth of language analyses between the participants with some kind of planning opportunities and those without. The participants who had an opportunity to plan were engaged in deeper thinking, such as linking certain knowledge with other knowledge and comparing two possible words to tell the story, no matter what types of planning they had. However, these behaviors were not observed for the participants who did not have an opportunity to plan.

In the next chapter, the findings of this chapter and those of the previous chapter (the results of L2 the speech data analyses) will be discussed.
CHAPTER 6 DISCUSSION

6.1 Introduction

This study has two primary purposes: (a) to investigate how different types of planning affect language learners’ performance in a narrative task, and (b) to explore what language learners do to ease the limitation of attention to produce more proficient speech. For these purposes, the following three research questions were posed:

1. How does time pressure affect the L2 learners’ oral production in a narrative task?
2. Are there any differences in the way pre-task and on line planning affect the L2 learners’ oral production in a narrative task, controlling for pre-existing Japanese proficiency differences?
   2.1 How do the different types of planning affect the fluency of L2 learners’ oral production in a narrative task?
   2.2 How do the different types of planning affect the complexity of L2 learners’ oral production in a narrative task?
   2.3 How do the different types of planning affect the accuracy of L2 learners’ oral production in a narrative task?
   2.4 Are there any trade-off effects among the fluency, complexity, and accuracy of L2 learners’ oral production in a narrative task?
3. What do L2 learners do during planning?

In this chapter, the results of the study are discussed for each research question, and then pedagogical implications are proposed. The chapter concludes with study limitations and directions for future task planning research.
6.2 Discussion of Results

6.2.1 Discussion for research question 1: Effects of time pressure

Research question 1 asks how the time pressure affects L2 learners’ oral performance and was investigated as a foundation for research question 2; that is, to identify the way to properly implement pre-task and on-line planning in this study so that the effects of each planning type could be examined separately.

As discussed in chapter 2, Yuan and Ellis (2003) claim that one of the problems in past planning studies, particularly those with inconclusive results for the effects of pre-task planning on accuracy in L2 learners’ speech, is that the amount of on-line planning was not controlled during task performance (Foster & Skehan, 1996; Mehnert, 1998; Ortega 1999; Skehan & Foster, 1997; Wendel, 1997; Wigglesworth, 1997). The participants in these studies likely engaged in different amounts of planning during the task, which might have affected their oral production. By imposing a time limit on task performance, Yuan and Ellis (2003) attempted to restrict on-line planning.

In the current study, the L2 speeches of learners who performed the task with and without a time limit were compared to examine the effects of the time limit on L2 oral production. It was found that time pressure affected fluency, but not complexity or accuracy. These results indicate that the participants in the Time Pressure group spoke more fluently because they felt pressed due to the time limit for their task performance. Hulstijn and Hulstijn (1984) reported similar results. Their participants with time pressure spoke significantly faster than those without such pressure, but there was no difference between these two groups in terms of the accuracy of the two grammatical patterns that were examined in their study.

In Yuan and Ellis (2003), the effects of time pressure were uncertain due to the lack of comparison between the participants with and without a time limit during the
story-telling task under the same planning conditions. However, considering the results of this study as well as the study by Hulstijn and Hulstijn (1984), the provision of time pressure may not be an appropriate means to control on-line planning when aiming to investigate the effects of pre-task planning on L2 oral performance. L2 learners may speak faster not because they have benefited from the opportunity of pre-task planning, but because they are hurrying to complete the task within the time limit. Following Yuan and Ellis (2003), it is possible that the participants in the No Time Pressure group in this study were engaged in more on-line planning than those in the Time Pressure group, but not to the extent that the complexity and accuracy of their production were affected.

To gain additional insights regarding the effects of time pressure, the participants who performed the narrative task with a time limit were asked during the interview how they felt about the time limit and how it affected their stories. A total of eight of the 50 participants with time pressure, half of whom had a pre-task planning opportunity, were interviewed. It was found that only half of the eight participants (2 participants each with and without pre-task planning) responded that they were conscious of the time limit during the task, although the speeches of the participants in the Time Pressure group were significantly faster than those produced without time pressure. Individual differences in reactions to time pressure were likely involved in the result due to the small number of participants who were interviewed. Thus, it is possible that not all of the participants with a time limit consciously hurried to complete the task. However, most of the participants in the Time Pressure group (7 of the 8 participants) reported that either the grammar or content of their stories might have been affected by the time limit. It is particularly important that three of the four participants with pre-task planning mentioned that the content of their stories had to be modified to finish the task within the time limit. This directly contradicts one of the primary purposes of providing an opportunity to plan, which is to allow L2 learners to access their knowledge without overloading their working memory. If the learners simplified their stories to adjust to the time limit, they
might have not attempted to create a story of the finest quality by taking advantage of the full extent of the linguistic knowledge available to them. Furthermore, the time pressure might have limited not only the content of the narrative stories, but also the amount of the stories. On average, the Time Pressure group produced only 260.46 moras and 238.82 meaningful moras, whereas the No Time Pressure group produced 309.54 moras and 271.92 meaningful moras.

Yuan and Ellis (2003) argue that providing time pressure reduces the amount of on-line planning in which participants can engage during task performance, and this would make it possible to examine the effects of pre-task planning more efficiently. However, the comments from the participants in this study indicate that there is a possibility that the time pressure is not only limiting the amount of on-line planning, but is also limiting the benefit of pre-task planning. This finding presents another drawback of time pressure, as well as the problem of rushing learners, which is that it affects their fluency and makes it difficult to examine the specific effect of pre-task planning on fluency.

6.2.2 Discussion for research question 2:

Effects of planning

Research question 2 addresses the effects of planning on L2 oral production and asks whether pre-task and on-line planning affect L2 oral production differently. Performance under four task conditions was compared to answer the question: no planning (G1), pre-task planning only (G3), on-line planning only (G5), and both pre-task and on-line planning (G6). Four sub-questions were included in research question 2. The first three sub-questions concern different aspects of oral production: fluency, complexity, and accuracy of the speech. Based on the factor analysis described in chapter 4, five measures were selected to examine the data: speech rate (fluency), number of clauses per T-unit (syntactic complexity), type-token ratio (TTR, lexical complexity), percentage of
error-free clauses (global accuracy), and percentage of correct particle use (specific accuracy). The last sub-question for research question 2 deals with trade-off effects among these three aspects. This section discusses the results for these four sub-questions.

6.2.2.1 The effects of task planning on the fluency of L2 oral production

The first sub-question for research question 2 asks how different types of planning affect the fluency of L2 speech. Previous planning studies have generally found that pre-task planning improves fluency (Foster & Skehan, 1996; Kawauchi, 2005a, 2005b; Mehnert, 1998; Ortega 1999; Sangarun, 2005; Skehan & Foster, 1997; Wendel, 1997; Wigglesworth, 1997; Yuan & Ellis, 2003), whereas on-line planning tends to decrease it (Yuan & Ellis, 2003). Speech rate was employed for this study to measure fluency and, unlike the results from these past studies, the results for this study revealed no statistically significant differences across planning conditions. The opportunity to plan prior to the task conferred no benefit to the participants in terms of fluency, nor did their fluency suffer following on-line planning.

One possible explanation for this result is the amount of on-line planning for the participants in G1 (no planning) and G3 (pre-task planning only). Unlike the participants in G5 and G6, the participants in G1 and G3 were not instructed to plan on-line. However, the stimulated recall protocols from four participants each in G1 and G3 revealed that they were actively engaged in on-line planning during the task performance, which might have decreased their fluency to the level that it was no different from the fluency of the participants in the on-line planning groups. For example, when asked about a pause while working on the last picture for the task, G1–3 answered as follows:

(11) G1–3: I was trying to figure out the word that…what was “to fix a problem,” cause I knew it before, but I forgot it.
Even the participants in G3, who had 10 minutes to think about the story before performing the task, still thought about what to say during the task performance as seen in the example below.

(12) G3–3: I was kind of wondering whether I should use

*oshieru* [to tell] as opposed to *setsu* [*setsumei*, to explain].

The majority of the participants in G1 and G3 also monitored their discourse as they spoke, which is another indication of on-line planning. G3–2 described his thinking during one of the pauses during his task performance in the following way:

(13) G3–2: I caught myself in saying *sorede* [and then] for the third time, so I was like, “Oh, I should use something else.”

Ortega (1999, 2005) and Kawauchi (2005a) reported that some of their participants, too, actively planned during the task, though they were allowed to plan for 10 minutes before the task just like the participants in G3 in this study. Based on these behaviors during the task performance, Ortega (1999) suggested that the opportunity to plan before the task eased up the strain on attention during the task and enabled the participants to work on their stories even more while performing the task. In other words, pre-task planning sets off on-line planning because additional attentional resources become available due to the provision of pre-task planning. With the extra attentional resources, Kawauchi (2005a) observed that learners were able to attend to parts of their
stories during the task performance to which they did not pay attention during pre-task planning.

Another possible explanation for active on-line planning by pre-task planners may be pertinent to an issue of memory (Kawauchi, 2005a; Wigglesworth & Elder, 2010). Kawauchi (2005a) reported that a few of the participants in her study described how they forgot what they had planned during the pre-task planning and had to make up the stories as they talked. Although the participants in Kawauchi’s (2005a) study who reported the problem of memory were at a low-intermediate level, a similar problem was mentioned by some of the participants in this study whose proficiency level in Japanese was at intermediate or above. The following excerpt from the interview with G3–2 illustrates this point. When being asked about a pause while working on the fifth picture where the main character was embarrassed because the solution for the problem was simple, he answered:

(14) G3–2: That was hazukashii [to be embarrassed].

(R: OK, that was what you were thinking.)

I was trying to remember, but it just slipped my mind.

Based on the comments from their study participants, Wigglesworth and Elder (2010) also pointed out the possibility that the restriction on memory might have resulted in the limited benefits of pre-task planning in their study. They explained that the advantage of pre-task planning could be realized only at the beginning of the task performance due to the memory problem.

Participants’ proficiency levels may have affected the non-significant results for fluency. The participants in this study were at intermediate to high-intermediate levels,
whereas the proficiency levels of the participants in the majority of the past planning studies were intermediate or lower (Ellis, 2009). Although these studies reported improved fluency as a result of pre-task planning, learners with different proficiency levels may have benefited in various ways, as the studies by Ortega (2005) and Kawauchi (2005a, 2005b) have shown. Kawauchi (2005b) compared the effects of planning on L2 oral production among learners at three different proficiency levels: low-intermediate, high-intermediate, and advanced. She found that of the three groups, the high-intermediate learners benefited the most in terms of speech rate and the percentage of repetition. This study also examined the effects of pre-task planning for the high-intermediate learners, but the results showed no positive effect of planning. One possible reason for the different outcomes between Kawauchi (2005b) and this study is that the high-intermediate learners in these studies were, in fact, not at a comparable proficiency level. In Kawauchi (2005b), proficiency levels were determined based on participants’ L2 learning experiences and TOEFL scores, whereas the proficiency levels in this study were based on the course levels in which the participants were enrolled. Ellis (2009) pointed out the lack of detailed descriptions of learner proficiency, which prevents researchers from understanding specific effects of planning for each proficiency level, as one of the limitations in planning studies.

Another explanation for the non-significant results for this study is that the narrative task that was used for this study might have been too easy for the participants to take advantage from pre-task planning. As a result, no planners were able to speak as fast as pre-task planners in this study. For the same reason, on-line planners might have not needed to take much time to think about the stories while performing the task and, thus, their fluency did not decrease comparing with no planners.
6.2.2.2 The effects of task planning on the complexity of L2 oral production

The second sub-question for research question 2 focuses on the complexity of L2 oral production. Two measures were used to evaluate the complexity of the participants’ oral production. The number of clauses per T-unit operationalized syntactic complexity, and type/token ratio (TTR) represented lexical complexity. Many previous studies found that syntactic complexity improves when L2 learners have an opportunity to plan before the task (Crookes, 1989; Foster & Skehan, 1996; Kawauchi, 2005a, 2005b; Mehnert, 1998; Ortega 1999; Sangarun, 2005; Skehan & Foster, 1997; Wendel, 1997; Wigglesworth, 1997; Yuan & Ellis, 2003) because they have more attentional resources available during task performance. For the same reason, generally pre-task planning has also been found to improve lexical complexity (Crookes, 1989; Gilabert, 2007; Tajima, 2003). Regarding on-line planning, Yuan and Ellis (2003) reported that on-line planning increased the syntactic complexity, but not the lexical complexity, of the resulting oral performance.

First, in terms of syntactic complexity, the results of this study did not show significant differences between the No Pre-Task Planning group and the Pre-Task Planning group, or between the No On-line Planning group and the On-line Planning group. Therefore, neither pre-task nor on-line planning helped the participants achieve enhanced complexity in syntax for their oral production in this study.

One of the possible reasons for these results may be task difficulty. As mentioned for the results of fluency, the task for this study may have been too easy for the participants to benefit from both pre-task and on-line planning. Due to the low difficulty, it is possible that even participants with no planning opportunity were able to produce stories as complex as planners.

Some studies that did not find a favorable effect of planning on syntactic complexity pointed to possible task effects in their results (Elder & Iwashita, 2005,
Iwashita, McNamara, & Elder, 2001; Tajima, 2003, Wigglesworth & Elder, 2010). In the case of Iwashita, McNamara, and Elder (2001), some raters in their study commented that the narrative task was not practical in terms of producing complex structures. To evaluate whether or not a task is too simple to generate complex structures, Tajima (2003) suggested gathering native speaker speech samples as baseline data. If the data from the native speakers show low syntactic complexity, then the task may be too simple to promote complex structures among learners as well.

In connection with tasks, Elder and Iwashita (2005) and Wigglesworth and Elder (2010) stated that narrative tasks may lead the participant to focus on the sequences of events in stories, rather than encourage more complex sentences. Based on the comments on the questionnaire that were written by the participants, Elder and Iwashita (2005) explained that the participants seemed to take the opportunity to plan as the time to work on the content of the stories, not grammatical items. A similar trend was observed in Wigglesworth and Elder (2010). They also analyzed the participants’ responses on the questionnaire about their strategy use during the pre-task planning for a narrative task and found that the participants reported attending to the meaning of the stories rather than to linguistic features.

In the present study, the participants were asked about their thinking processes during planning and task performance during the retrospective interviews. Most of the time, they answered that they were thinking about the content of the stories, and they seldom mentioned grammar. On average, the frequency that the participants talked something about grammar ranged from 0.88 times per interview for the Pre-Task Planning group (G3 and G6) and the No On-line Planning group (G1 and G3) to 1.13 times for the No Pre-Task Planning group (G1 and G5) for participants who were interviewed in each planning group. The On-line Planning group (G5 and G6) mentioned about grammar once on average during the interviews.
With regard to lexical complexity, it was found that the participants who had an opportunity to plan before the task produced stories with significantly less variety of vocabulary, which was opposite of what was expected. On-line planning did not affect lexical complexity. This study selected type-token ratio (TTR) based on the results from the factor analysis that was conducted prior to the speech data analysis. As described in chapter 4, for lexical complexity measures TTR and modified TTR (type/√2*token) were examined, and it was found that TTR loaded highly on the factor that seemed to indicate lexical complexity and lexical accuracy, whereas the modified TTR loaded highly on the factor that seemed to concern syntactic complexity and syntactic accuracy. Correlation analysis further confirmed that TTR was independent from syntactic complexity measures whereas the modified TTR was not. However, it has been pointed out that TTR may not be reliable when the length of the samples varies (Tajima, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998). Wolfe-Quintero, Inagaki, and Kim (1998) explain that one of the issues with TTR is that it tends to decrease as production gets larger because the same words are likely to be used repeatedly.

To compare the amount of oral production between the Pre-Task Planning group (G3 and G6, PG2) and the No Pre-Task Planning group (G1 and G5, NPG2), the average number of moras for each group was calculated. Meaningful moras, which are the number of moras that exclude any redundant moras such as repeated or replaced words, were also calculated to evaluate the actual production of a story. For an added comparison, moras and meaningful moras for the On-line Planning group (G5 and G6, OLP) and the No On-line Planning group (G1 and G3, NOLP) were also calculated. Tables 54 and 55 display the descriptive statistics for the number of moras and the number of meaningful moras by group respectively.
Table 54. Descriptive statistics for the number of moras by group

<table>
<thead>
<tr>
<th>Pre-task planning</th>
<th>On-line planning</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>− Mean (SD)</td>
<td>272.67 (125.36)</td>
<td>315.5 (111.35)</td>
<td>294.08 (119.27)</td>
</tr>
<tr>
<td>+ Mean (SD)</td>
<td>346.42 (123.5)</td>
<td>379.62 (117.6)</td>
<td>361.91 (120.58)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>309.54 (128.62)</td>
<td>345.42 (117.53)</td>
<td></td>
</tr>
<tr>
<td>Adjusted mean</td>
<td>311.78</td>
<td>344.96</td>
<td></td>
</tr>
</tbody>
</table>

The results show that for both total moras and meaningful moras, the PG2 spoke more (361.91 moras and 328.27 moras, respectively) than the NPG2 (294.08 moras and 251.79 moras, respectively). Comparing the NOLP group and the OLP group, the results

Table 55. Descriptive statistics for the number of meaningful moras by group

<table>
<thead>
<tr>
<th>Pre-task planning</th>
<th>On-line planning</th>
<th>Mean (SD)</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>− Mean (SD)</td>
<td>233.96 (98.26)</td>
<td>269.63 (106.96)</td>
<td>251.79 (103.19)</td>
</tr>
<tr>
<td>+ Mean (SD)</td>
<td>309.88 (104.99)</td>
<td>349.29 (99.23)</td>
<td>328.27 (103.11)</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>271.92 (107.66)</td>
<td>306.80 (109.87)</td>
<td></td>
</tr>
<tr>
<td>Adjusted mean</td>
<td>273.97</td>
<td>307.07</td>
<td></td>
</tr>
</tbody>
</table>
show that the OLP group spoke more than the NOLP group at a total moras of 344.96 and a meaningful moras of 307.07, respectively, whereas the NOLP group produced 309.54 moras in total and 271.92 meaningful moras. The results revealed that the two groups with an opportunity to plan produced stories that were longer than the groups without planning either before or during the task.

These data were submitted to a two-way MANCOVA in which pre-task planning and on-line planning were independent variables, and SPOT scores were covariates. The alpha level was set at .05. The results of Wilks’ lambda are shown in Table 56.

Table 56. Wilks’ lambda for a two-way MANCOVA for pre-task planning and on-line planning effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>.772</td>
<td>12.859</td>
<td>2</td>
<td>87</td>
<td>.000**</td>
</tr>
<tr>
<td>On-line planning</td>
<td>.965</td>
<td>1.598</td>
<td>2</td>
<td>87</td>
<td>.208</td>
</tr>
<tr>
<td>Pre-task planning × On-line planning</td>
<td>.983</td>
<td>.746</td>
<td>2</td>
<td>87</td>
<td>.477</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

The results of the Wilks’ lambda test showed that there was a significant main effect for pre-task planning ($F = 12.859, p < .01$), but not for on-line planning ($F = 1.598, p = .208$). The interaction effect between pre-task and on-line planning was also not significant ($F = .746, p = .477$). Therefore, only the significant results of the follow-up ANCOVA for pre-task planning were considered.
Tables 57 and 58 display the results of the follow-up ANCOVAs for moras and meaningful moras. The covariates were significant for both moras and meaningful moras.

Table 57. Follow-up ANCOVA for task planning conditions on moras

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>101354.803</td>
<td>7.433</td>
<td>.008*</td>
</tr>
<tr>
<td>On-line planning</td>
<td>25323.020</td>
<td>1.857</td>
<td>.176</td>
</tr>
<tr>
<td>Pre-task planning × On-line planning</td>
<td>1724.502</td>
<td>.126</td>
<td>.723</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Table 58. Follow-up ANCOVA for task planning conditions on meaningful moras

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task planning</td>
<td>131049.475</td>
<td>13.206</td>
<td>.000**</td>
</tr>
<tr>
<td>On-line planning</td>
<td>25196.930</td>
<td>2.539</td>
<td>.115</td>
</tr>
<tr>
<td>Pre-task planning × On-line planning</td>
<td>62.431</td>
<td>.006</td>
<td>.937</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

The results of the follow-up ANCOVA revealed that PG2 produced significantly longer narratives as identified by both moras and meaningful moras. Considering these results, it may be possible that the significantly smaller TTR for PG2 compared to NPG2 was due to the larger sample size that PG2 produced.
Ortega (1999) pointed out another methodological issue in using TTR. She reported that no significant difference in TTR was found between performance with or without a pre-task planning opportunity. She explained that there was a possibility that disfluencies, such as self-corrections and false starts, might have resulted in increased lexical types for non-planners. Ortega (1999) claimed that pre-task planners were more likely able to solve these disfluency problems during planning, thus resulting in fewer unnecessary words in their task performance.

Additionally, trade-off effects may have had something to do with the results for lexical complexity in this study. TTR was involved with trade-off effects for G3 and G5. In the case of G3, who had an opportunity to plan before the task but not during the task, and G5, who had no opportunity to plan before the task but did plan during the task, TTR was negatively correlated with the percentage of error-free clauses. Furthermore, the results for G5 showed that TTR was negatively correlated with speech rate as well. These results suggest that there is complex interaction between lexical complexity and other aspect of oral production. Trade-off effects that were observed in this study are discussed below (section 6.2.2.4).

6.2.2.3 The effects of task planning on the accuracy of L2 oral production

The third sub-question for research question 2 asks how different types of planning affect the accuracy of L2 oral production. Two measures were used to calculate the accuracy of L2 oral production: the percentage of error-free clauses (global measure) and the percentage of correct particle use (specific measure). Unlike fluency and complexity, for which past studies have generally obtained positive results of planning, the effects of planning on accuracy have been inconclusive, because some studies have shown positive effects whereas other studies have shown partial or no effect (Ellis, 2005,
The results for this study showed no significant effect of pre-task planning or on-line planning for either global or specific measures.

A purpose of including specific accuracy measures is to detect any small changes in L2 production that global measures may fail to depict (Ortega, 1999). It is possible that some specific accuracy measures show significant results whereas others do not when examining the same data as shown in Ortega (1999). In the present study, planning did not assist in improving accuracy of particles, but there is a possibility that some other forms, particularly the forms that even advanced learners generally have difficulty to produce correctly, may have had a positive effect of planning. Different results might have been obtained if other specific measures would have been used.

Another possible explanation for non-significant effect on both specific and global accuracy is the task type, as discussed in the previous section regarding syntactic complexity. This study used a narrative task and, consequently, participants’ focus may have been on content of their stories, rather than on linguistic form (Elder & Iwashita, 2005), which may have resulted in no significant effects on accuracy or on syntactic complexity. During the interviews, a majority of the participants mentioned something about form, such as that they were thinking about what vocabulary or grammar to use to describe particular parts of the stories, but that does not necessarily mean that they paid attention specifically to accuracy of form. For example, G3−1 reported as follows when watching himself in the video working on the fourth picture for the task during pre-task planning:

(15) G3−1: Here, I knew that I was going to say something-*nagara* [while], so first I was thinking what I should say here. And then, *nagara* [while], and then, what he’s doing.
During the task performance, G3–1 described the fourth picture as follows. The part where G3–1 used *nagara* is underlined.

(16) *Sensee to Hiroshi-kun ga konpuuta wo dooshita no ka wo kangae nagara, tonari no gakusee wa koodo ga tsui, tsunagaru to iundesu ka. Sono gakusee ga koodo wo, koodo wo mirun desu.*

[While the teacher and Hiroshi-kun (the main character) were figuring out what happened to the computer, the student next to Hiroshi, the cord is connected. The student saw the cord.]

G3–1 explained that he wanted to use *nagara* [while] to describe the fourth picture where a student next to Hiroshi (the main character) saw the cord was disconnected as Hiroshi and the teacher were checking the computer. However, using *nagara* in this sentence is incorrect because the subjects for the two clauses that are connected by *nagara* have to be the same. The subjects cannot be different, such as Hiroshi and the teacher for the subordinate clause and the student for the main clause, as shown in (16). Although G3–1 talked about how he wanted to use *nagara*, it was not clear from the interview whether he considered the accurate use of *nagara* for this particular sentence.

Relevant to attention to accuracy, Ortega (1999, 2005) reported that some of her participants were oriented towards grammatical accuracy, whereas others were oriented towards communication. She explained that the latter group tended to be more accepting of their errors and seemed even to consider that being accurate was not practical for L2
learners. In the current study, no clear distinction was observed between participants with more grammatical versus more communicative mindsets. However, there were some cases in which participants reported that they prioritized communicative needs over accuracy. Consider the following comment from G6–2, who remarked that he was aware of his mistakes, but did not correct them:

(17) G6–2: There were a few points where I made mistakes and thought about them, but didn’t change them. I just went on in the story.

G6–2 was encouraged to plan on-line in addition to having an opportunity to plan before the task, and he understood that he could spend as much time as he wished to tell the story. Yet, he decided not to correct his mistakes even though he was aware of them, but kept on performing the task. As the example above shows, it is possible that some participants with planning opportunities in this study valued communicative needs over accuracy, and did not take advantage of their available attentional resources during the task.

Another possible reason for not finding significant differences in accuracy across planning conditions may be that the participants were not sure how to make the most of their planning opportunities. In this study, there was no instruction as to what and how to plan for the task, and it was completely up to the participants to decide what to do during their planning time. It is possible that the participants were not able to take full advantage of an opportunity to plan because they did not know what to do besides doing the same things that they would have done had they not have planning opportunities. In contrast, some studies that involved guided planning reported results showing improved accuracy (Mochizuki & Ortega, 2008, Sangaran, 2005). Mochizuki and Ortega (2008) compared
no planning, unguided pre-task planning, and guided pre-task planning conditions regarding the use of relative clauses. They found that the guided planning group was more accurate in their use of relative clauses, whereas there was no difference between the no planning group and the unguided planning group. Mochizuki and Ortega (2008) suggest that the provision of planning without guidance may not necessarily encourage learners to focus on form. The participants in Mochizuki and Ortega (2008) were at an elementary level, and the researchers suggested that L2 learners with limited proficiency may benefit from guided planning, but more proficient learners, such as the participants in the current study, may also be assisted by more specific instruction on how to plan for the task. The trend of these behaviors for planning across planning conditions, without taking advantage of planning opportunities, was also observed in the retrospective interviews for this study, which was reported in the previous chapter and will be discussed in a later section about the use of strategies.

6.2.2.4 Trade-off effects among the fluency, complexity, and accuracy of L2 oral production

Previous studies have reported trade-off effects between complexity and accuracy (Skehan & Foster, 1997), fluency and accuracy (Wendel, 1997; Yuan & Ellis, 2003), and lexical complexity and grammatical accuracy (Yuan & Ellis, 2003). In this study, a trade-off effect was observed between lexical complexity and global accuracy when the combined data for G1, G3, G5, and G6 (r = −.211, p < .05) were examined, which generally echoed the study by Yuan and Ellis (2003), who claimed trade-off effects between these two aspects. They found that the pre-task planners had higher lexical complexity, but lower grammatical accuracy (global and specific), whereas the on-line planners showed more accurate production (global and specific), but less lexical complexity. This study further confirmed their claim, at least for global accuracy, by obtaining statistically significant negative correlation between lexical complexity and
global accuracy. A comparison within each planning group also found a trade-off effect for G3 at a significant level ($r = -0.527, p < .01$) and for G5 at a near-significant level ($r = -0.357, p = .087$). These results suggest that L2 learners tend to pay attention to either lexical complexity or accuracy in general, but particularly when they have an opportunity to plan before or during task performance. Interestingly, lexical complexity and global accuracy were positively correlated ($r = 0.484, p < .01$) when the participants were allowed to plan both before and during the task. It seems that having an opportunity to plan either before or during the task is not enough to lessen L2 learners’ cognitive load to enable them to pay attention to various aspects of language, and they still needed to prioritize lexical complexity over accuracy or vice versa. L2 learners can allocate their attention to both lexical complexity and accuracy only when they have time to plan before the task and in addition, are able to engage in planning during the task.

A negative correlation was found between fluency and lexical complexity for G5 (on-line planning without pre-task planning, $r = -0.442, p < .05$), which suggests another trade-off effect. This trend was found only for G5, and not for any other group, including the group that combined all of the participants across all planning conditions (Tables 42–46). The data for the G6 group, who also planned on-line, did not show the same trade-off effect between fluency and lexical complexity as G5 did. In the case of G6, they were allowed to plan before the task and were free to access their lexical knowledge during that time. Because of the pre-task planning opportunity, it seems that the participants in G6 no longer needed to take time during the task to enrich the variety of vocabulary in their stories. However, it cannot be assumed that the pre-task planners in this study (G3 and G6) planned for vocabulary during the pre-task planning because TTR in their oral production was significantly less than that of non pre-task planners. Rather, it appears that the focus of pre-task planning was the content of the stories, considering the result that their stories were longer than the other groups.
6.2.3 Discussion for research question 3: Strategy use

Research question 3 asks what strategies L2 learners use during task planning. To answer this question, strategy use was compared among three groups: the Pre-Task Planning group (G3, G4, and G6), the On-line Planning group (G5, and G6), and the No On-line Planning group (G1, G2, G3, and G4). The results showed that all three groups used a greater variety of metacognitive strategies (Pre-Task: 3.83 types, On-line: 4 types, No On-line: 3.63 types) than cognitive strategies (Pre-Task: 3.33 types, On-line: 1.88 types, No On-line: 1.69 types) or social/affective strategies (Pre-Task: 0.17 types, On-line: 0.75 types, No On-line: 0.13 types). These results did not replicate the results of Ortega (2005), which examined planning processes for pre-task planning and reported that the participants used metacognitive and cognitive strategies to a similar extent (5.20 types and 5.91 types respectively). The inconsistency between the results in her study and this study regarding pre-task planners’ strategy use may be due to the differences in the design of these studies. In this study, participants were instructed to tell their stories based on the pictures as if they were talking to someone who had never seen these pictures, but in reality they did not have anyone at present besides the researcher to listen to their stories during task performance. On the other hand, participants in Ortega (2005) were paired, and one told a story to the other, who needed to put the pictures about the story in order as he or she listened to the speaker and later wrote about the story that the partner had recounted. Because of this communicative setting to perform the task, Ortega (2005) reported that many speakers were sensitive to their listener’s need to understand the story and paid attention to their choice of language, particularly vocabulary, which might have resulted in increased use of lexical compensation strategies, such as Approximating and Circumlocution and synonyms. In her study, 50% of participants used Approximation and 43% used Circumlocution and synonyms, whereas in this study 33.33% used Approximation and 0% used Circumlocution and synonyms.
A limited use of cognitive strategy is particularly evident during task performance regardless of the opportunity for on-line planning (On-line: 1.88 types, No On-line: 1.69 types). One of the reasons for this result is that three of the strategies were writing-related: Writing for retrieval, Writing for later recall, and Highlight and postpone. Note taking was not available during the task, though it was during pre-task planning, and therefore, none of the participants used these three strategies while performing the task. Another possible reason may have originated from the nature of on-line planning, which Yuan and Ellis (2003) characterize as careful speech planning and monitoring. In fact, more than half of the participants in both the On-line Planning and No On-line Planning groups were engaged in Organizational planning (100% of the On-line Planning group and 75% of the No On-line Planning group) and Self-monitoring (75% and 93.75% respectively) while performing the task. These results suggest that L2 learners’ behaviors may be geared to metacognitive strategy use rather than cognitive strategy use during the task.

The results of this study revealed that all of the groups used certain strategies, such as Organizational planning and Problem identification, but did not use other strategies. The opportunity to plan, whether pre-task or on-line, did not necessarily encourage the participants to use a greater variety of strategies. It is noteworthy that even with 10 minutes to plan for the task, pre-task planners did not increase the overall number of strategy types they used. Wigglesworth and Elder (2010) compared strategy use when participants had one minute and two minutes of pre-task planning time and reported that the participants with two minutes to plan used significantly more strategies than those with one minute of planning time although the difference was small. If just one minute of difference in planning time makes such a difference, why were the pre-task planners in this study unable to use a greater variety of strategies during the 10-minute pre-task planning period? One possible reason may be that the task in this study was not challenging enough to encourage participants to use strategies. The possibility that task
difficulty may have resulted in non-significant effects of any types of planning on fluency, complexity, and accuracy of L2 production was discussed in sections 6.2.2.1 through 6.2.2.3 above. Similarly, the task difficulty may have played a role in how participants in this study selected strategies.

A limited use of strategies corroborates what was discussed in section 6.2.2.3, which is the possibility that participants may have not known what to do with the task planning opportunity and, therefore, may not have been able to maximize its potential benefits. They may have simply repeated what they would do when no planning opportunities were available. Similar observation was made in some studies such as Elder and Iwashita (2005) and Mochizuki and Ortega (2008), and these results suggest that there is a need for training in strategy use during planning time (Sangarun, 2005), so that learners deploy a variety of strategies to help them produce more proficient speech.

Although it appears that there is little quantitative difference in strategy use across planning conditions, it was pointed out that planners, whether pre-task or on-line, seemed to engage in deeper language processing while preparing for the task. It may be the case that the provision of planning reduced attentional load and allowed learners to access their knowledge. Qualitative differences in strategy use between high and low proficiency learners have been observed in Chamot (2001) and Ortega (2005), but it seems that these differences are present for such a temporal condition like provision of a planning opportunity.

6.2.4 Summary of discussion

The results of this study suggest the need for reconsideration of providing time pressure to examine the effects of pre-task planning. The provision of time pressure resulted in faster speech regardless of the opportunity of pre-task planning, which in turn affected the results for pre-task planning. Furthermore, there is a possibility that the participants altered their stories in terms of the grammar if they were no-planners and in
terms of content if they were pre-task planners. Considering these issues, other approaches to investigate the effects of pre-task planning without the effects of on-line planning should be sought out.

As regards the effects of different types of planning, no significant effects on fluency, complexity, and accuracy in L2 oral production were found for any planning conditions in this study, except for the negative effect of pre-task planning on lexical complexity. There are several possible explanations for these unexpected results, which can be generally categorized in two factors: learner and task variables, both of which are part of the framework that Ellis (2009) proposes for investigating task planning. One of the learner variables that possibly played a role in the results for this study was the participants’ familiarity with task planning. They may have not been able to make good use of planning because they simply did not know what to do to plan due to unfamiliarity with speaking under planning conditions.

Turning to task variables, the narrative task for this study may have not been challenging enough for the participants to benefit from the planning opportunities. Another task variable that could have been the cause of the results is the nature of the narrative task; that is, participants’ focus during planning might have been directed to the sequence of the stories rather than forms, which could have affected the complexity and accuracy of their production in particular.

Two sets of trade-off effects were found in this study. The first set was lexical complexity and global accuracy, which was found in production for the pre-task planning only and the on-line planning only groups. These results indicate that the provision of pre-task and on-line planning individually is not sufficient to enable L2 learners to pay attention to both accuracy and complexity. The other set was fluency and lexical complexity for the on-line planning only group, suggesting that pre-task planning may assist learners in rapid retrieval of lexical items.
Regarding strategies for planning, it was generally found that planning opportunities did not encourage the use of a greater variety of strategies. Some of the possible reasons for this result are comparable with what was discussed for the results for research question 2. One is the difficulty level of tasks, and the other is inadequate use of planning opportunity. Although there were no apparent differences in the use of strategies across planning conditions, the results revealed a qualitative difference in processes of planning between planners and non-planners. More research should be conducted to advance our understanding of the effects of planning on more proficient learners’ oral production as well as on their planning processes, but there are some pedagogical implications that this study can offer, which will be discussed in the following section.

6.3 Pedagogical Implications

The results of this study suggest that neither pre-task planning nor on-line planning is effective for intermediate to high-intermediate L2 learners to improve task performance in terms of fluency complexity, and accuracy. However, with a pre-task planning opportunity, it was found that the participants spoke more than those without pre-task planning. From the perspective of the Output Hypothesis, this finding itself offers a reason to include pre-task planning in L2 classroom activities because, simply stated, more output means more opportunities to notice something about the language, test language-relating hypotheses, and think about the language: three functions of output in facilitation of language learning (Swain, 1995, 1998, 2005). In addition, those participants who were allowed to plan for the task, whether pre-task or on-line, seemed to engage in deeper language processing, which is linked to improvement in longer-term memory (Swain, 2005) and, therefore, learning. Even without the positive effects of planning on oral production, which are determined by quantitative measures, it is worth providing L2 learners with opportunities for planning, of whichever types fit the frameworks of the class.
When implementing task planning in L2 classroom activities, task and individual learner variables should be taken into account to ensure that task planning will contribute to learning. As discussed earlier one of the possible reasons for the non-significant results of this study was the difficulty of the task relevant to L2 learners’ proficiency levels. If learners are not forced to reach to upper limits of their internal knowledge to prepare for a task, it is possible that the provision of planning may be of little use to improve the quality of their production. Careful assessment of task difficulty that is suitable for particular L2 learners’ proficiency levels is necessary for learners to benefit from task planning opportunities.

The results of the present study showed that the participants across planning groups used a limited variety of strategies, which suggests the possibility that L2 learners are not able to fully utilize the opportunity to plan for the task regardless of the types of planning they are given. Learners may not know how best to take advantage of the planning opportunities to produce speech at an optimal level since they are not used to performing a task under planned conditions (Elder & Iwashita, 2005). Thus, it is necessary to consider how to enable L2 learners to plan more effectively. One way of responding to this issue, a more immediate and direct way, would be to provide specific guidance for planning. Mochizuki and Ortega (2008) and Sangarun (2005) reported that they were able to lead their participants to plan in certain ways by providing guidance during pre-task planning, which resulted in more proficient oral production. However, these studies concerned only pre-task planning, and no research has been conducted for the effects of guidance for on-line planning. Therefore, guided planning can be considered only for pre-task planning.

Another possible solution for non-efficient use of planning, a more long-term and continuing approach, may be to provide strategy training on how to plan for a task. In general, past studies have reported a positive effect of strategy training on speaking (Chamot, 2001; Cohen, Weaver, & Li, 1998). Although the focus of these studies was not
specifically on planning it is worth considering strategy training so that the learners become able to plan to perform a task more efficiently (Elder & Iwashita, 2005; Wigglesworth & Elder, 2010).

6.4 Limitations of the Present Study

All research studies have limitations, and the present study is no exception. First, this study targeted Japanese learners at intermediate and high-intermediate levels. However, some studies, such as Kawauchi (2005a) and Ortega (2005), reported that L2 learners with different proficiency levels seemed to use different strategies. Kawauchi (2005b) also reported that the opportunity to plan affected L2 oral production differently depending on the learners’ proficiency levels. Therefore, the findings of this study cannot be generalized to L2 learners across proficiency levels.

Second, the speech data in this study were analyzed for fluency, complexity, and accuracy to answer research question 2, which addresses the effects of task planning on L2 oral production. In this study, a factor analysis was used to determine the most appropriate measures to assess participants’ oral production, and five measures were chosen: number of moras per minute for fluency, number of clauses per T-unit for syntactic complexity, type-token ratio for lexical complexity, percent of error-free clauses for global accuracy, and percent of correct use of particles for specific accuracy. Although there is no doubt that the most ideal measures were selected for this study based on the results of the factor analysis, their validity and reliability are still an issue, simply because there have not been many studies that evaluate measures for L2 oral production in Japanese (Iwashita, 2006; Tajima, 2003). It is possible that the results of this study would have been different if other measures had been employed.

Another limitation of this study concerning measures is that the present study concentrated on quantitative differences in participants’ oral production and did not include qualitative assessment. Therefore, any differences in the quality of the narratives
regarding the content were not taken into consideration. Tajima (2003) used subjective ratings to examine the effects of pre-task planning on the message of L2 oral production such as how well each task performance met the task requirements and how well the information was organized. She reported that the results revealed higher quality for planned speeches. Ortega (1999) also used subjective ratings to examine overall impression and quality of narratives to evaluate L2 oral production. Although the results did not show a significant effect of pre-task planning, it was found that the planned speeches were generally rated higher. In the current study, the amount of production was increased for the narratives with a pre-task planning opportunity, and therefore, it is likely that the planners included more information in their stories. However, it is unknown whether there were any changes in quality of oral production with regard to discourse organization and task accomplishment, for instance.

Third, although the speech data for this study were collected from a total of 143 participants, 21 to 25 participants each in six experimental groups, a small number of participants, four participants per group, was involved in retrospective interviews for strategy analyses. It is undeniable that the individual differences strongly affected the results of strategy analyses, which limits the generalizability of this study.

Fourth, to investigate what L2 learners do to prepare for a task, this study examined the types of strategies used, but not the frequency of strategy use. It was found that the participants tended to use or not to use similar strategies across planning conditions. However, some L2 learners may have used strategies more frequently than others depending on the given planning conditions (Kawauchi, 2005a). By examining the frequency of each strategy that participants used, more differences as a result of types of planning might have been found.

Finally, the present study attempted to investigate the effects of different types of planning on L2 oral production as well as to examine planning processes. Some studies suggest providing time pressure during task performance for no on-line planners so that
the amount of on-line planning could be kept to a minimum and would not affect the results for other planning conditions such as no planning and pre-task planning only (Elder & Iwashita, 2005; Wigglesworth & Elder, 2010; Yuan & Ellis, 2003). However, this study opted not to provide time pressure for no on-line planners because the results for research question 1 showed that time pressure might have hurried them to speak faster. Instead of providing time pressure for no on-line planners, on-line planners in this study were encouraged to think as they performed the task and practiced on-line planning prior to the data collection. On the other hand, no on-line planners received no instruction as to what they should do during the task. Therefore, the participants in G1 (no planning) and G3 (pre-task planning only) were free to engage in on-line planning, which may have affected the results for this study. While encouraging on-line planning to investigate its effects, it is also necessary to find a way to control the amount of on-line planning for participants in no planning and pre-task planning only conditions so that the effects of these task conditions can be examined without the concurrent effect of on-line planning. At the same time, a question arises whether it is necessary to investigate the specific roles of no planning or pre-task planning when on-line planning is involved in any communicative situations in both L1 and L2. It is more important to make sure that on-line planners spend ample time, more than usual, on planning while performing a task to investigate the effects of no planning, pre-task planning, on-line planning, and the combination of pre-task and on-line planning. Furthermore, a way to determine whether on-line planners actually plan during the task more than no on-line planners do should be considered (Ellis, 2009).

6.5 Directions for Future Research

Several future research directions can be proposed based on the results of this study. First, the results of this study showed that intermediate and high-intermediate L2 learners might not have benefited from planning opportunities. The typical learner
samples in previous planning studies are at the proficiency levels of intermediate and lower (Ellis, 2009), and therefore, it is not clear whether planning is as effective as what has been reported in the literature depending on L2 proficiency levels. Kawauchi (2005b) reported that high-intermediate learners generally benefited from pre-task planning more than intermediate and advanced learners, but more studies are necessary to determine how provision of planning affects oral production of L2 learners with different proficiency.

Another area that needs more research is L2 learners’ strategy use during planning, particularly different strategy use depending on the learners’ proficiency levels. As mentioned earlier, Kawauchi (2005a) and Ortega (2005) reported that there were differences in strategy use between intermediate and advanced learners. It is necessary to investigate the interaction between L2 proficiency level and strategy use to determine whether or not the selection of task planning strategies changes as learners’ L2 abilities develop.

As discussed in the previous section, there is an issue of validity and reliability of the measures for L2 oral production. Research is necessary to investigate more valid and reliable measures, particularly for less commonly taught languages like Japanese. In addition, holistic measures to examine the quality of L2 oral production are needed. These measures may be able to capture changes in task performance that cannot be identified by quantitative measures and could shed light on the effects of planning that may have not been discovered.

The current study involved only a narrative task, and research is needed to examine how planning affects different tasks. There are some prior studies that report differences in how task planning affected L2 learners’ oral production depending on the types of tasks they performed (Crookes, 1989; Foster and Skehan, 1996; Mehnert, 1998). Performance of some types of tasks may be affected more by the opportunity to plan than
others. It is also necessary to investigate processes of planning for other types of tasks to understand L2 learners’ planning processes better.

Although the results of this study did not show positive effects of pre-task or on-line planning on L2 oral production, many previous studies have obtained results that suggest beneficial roles of planning, particularly for fluency and complexity. One of the issues that still remains in task planning research is the lack of a longitudinal study that would examine whether improvement in a task performance that is observed for a learner under a particular planning condition will last and will impact his/her future oral production (Ellis, 2009). In other words, there need to be studies that investigate how planning can contribute to second language acquisition. Future research should address this very fundamental question by including studies that examine long-term effects of planning.

6.6 Conclusion

The present study explored the research questions that address the following three factors: (a) effects of time pressure on L2 oral production, (b) effects of pre-task and on-line planning on L2 oral production, and (c) processes of pre-task and on-line planning. Results of this study suggest that L2 learners may speak faster when they perform a task under time pressure. Thus, the provision of time pressure may disturb its primary purpose; that is, to investigate the sole effects of pre-task planning. The analyses of the effects of pre-task and on-line planning on Japanese learners’ oral production for a narrative task revealed no significant differences in task performance across planning conditions regarding fluency, complexity, and accuracy, except for lexical complexity with and without pre-task planning. Therefore, it appears that neither pre-task nor on-line planning assisted L2 learners in producing speeches with higher quality. However, the planners, particularly planners with a pre-task planning opportunity, produced longer narratives than no planners, which may facilitate more opportunities for learning. With
regard to planning processes, for the most part, it was found that L2 learners used similar strategies regardless of types of strategies that they were provided. Nevertheless, participants engaged in deeper language processes while planning for narratives, regardless of pre-task or on-line planning.

This study had important pedagogical implications and suggested useful future research directions. Task planning is one of the areas of second language acquisition research that connects classroom instruction with second language development and has been investigated with a great interest. It is hoped that the findings of this study will stimulate more research on task planning.
APPENDIX A: QUESTIONNAIRE

Name: _____________________________________
Address: _______________________________________________________
Phone: (     ) ____________________
E-mail: ___________________________
Native language: ___________________ Gender: _______ Age: _______
Major: _______________________

1. How long have you been studying Japanese?

__________________________________________________________________

2. Have you been to Japan?

   Yes    No
   ↓
   When, and for how long? What was the purpose?

__________________________________________________________________

3. Is anyone in your immediate family a native speaker of Japanese?

   Yes    No
   ↓
   What is the relationship? How often do you speak to the person in Japanese?

__________________________________________________________________

4. Do you speak Japanese regularly outside of your Japanese class?

   Yes    No
   ↓
   In what capacity, and how often?

__________________________________________________________________

5. Do you regularly engage in activities that involve speaking (speech, debate, etc.)?

   Yes    No
   ↓
   What activities, and how often?

__________________________________________________________________
Familiarization Task

Please look at the following pictures and describe in Japanese what happened to a man, Yoshio, who was going to see a movie at 3:00pm.
Main Task

Please look at the following pictures and describe in Japanese what happened to a college student, Hiroshi.
APPENDIX C: TASK INSTRUCTIONS

G1: Familiarization task
Thank you for participating in this study. This is a practice session. You will receive a set of pictures describing something that happened to Yoshio. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Yoshio to someone who has never seen the pictures.
You have 30 seconds to look through the pictures. There is no time limit to complete the task.

G1: Main task
You will receive a set of pictures describing something that happened to a college student, Hiroshi. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Hiroshi to someone who has never seen the pictures.
You have 30 seconds to look through the pictures. There is no time limit to complete the task.
Once again, thank you for your cooperation!

G2: Familiarization task
Thank you for participating in this study. This is a practice session. You will receive a set of pictures describing something that happened to Yoshio. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Yoshio to someone who has never seen the pictures.
You have 30 seconds to look through the pictures. You will have 1 minute and 30 seconds to tell the story.

G2: Main task
You will receive a set of pictures describing something that happened to a college student, Hiroshi. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Hiroshi to someone who has never seen the pictures.
You have 30 seconds to look through the pictures. You will have 1 minute and 30 seconds to tell the story.
Once again, thank you for your cooperation!

G3: Familiarization task
Thank you for participating in this study. This is a practice session. You will receive a set of pictures describing something that happened to Yoshio. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Yoshio to someone who has never seen the pictures.
You have 5 minutes to prepare for this task. You may take notes on the sheet of paper provided, but please do not write out everything you intend to say either in English or in Japanese. The paper will be removed at the end of the preparation time and you will not have it while you are speaking. There is no time limit to complete the task.
G3: Main task
Thank you for participating in this study. You will receive a set of pictures describing something that happened to a college student, Hiroshi. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Hiroshi to someone who has never seen the pictures.
You have 10 minutes to prepare for this task. You may take notes on the sheet of paper provided, but please do not write out everything you intend to say either in English or in Japanese. The paper will be removed at the end of the preparation time and you will not have it while you are speaking. There is no time limit to complete the task.
Once again, thank you for your cooperation!

G4: Familiarization task
Thank you for participating in this study. This is a practice session. You will receive a set of pictures describing something that happened to Yoshio. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Yoshio to someone who has never seen the pictures.
You have 5 minutes to prepare for this task. You may take notes on the sheet of paper provided, but please do not write out everything you intend to say either in English or in Japanese. The paper will be removed at the end of the preparation time and you will not have it while you are speaking. You will have 1 minute and 30 seconds to tell the story.

G4: Main task
Thank you for participating in this study. You will receive a set of pictures describing something that happened to a college student, Hiroshi. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Hiroshi to someone who has never seen the pictures.
You have 10 minutes to prepare for this task. You may take notes on the sheet of paper provided, but please do not write out everything you intend to say either in English or in Japanese. The paper will be removed at the end of the preparation time and you will not have it while you are speaking. You will have 1 minute and 30 seconds to tell the story.
Once again, thank you for your cooperation!

G5: Familiarization task
Thank you for participating in this study. This is a practice session. You will receive a set of pictures describing something that happened to Yoshio. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Yoshio to someone who has never seen the pictures.
You have 30 seconds to look through the pictures. There is no time limit to complete the task.
I encourage you to take time to think about the story while you are telling it. To remind you of doing so, I will tell you “don’t forget that you can take time to think about what you are going to say” during the task.
G5: Main task
Thank you for participating in this study. You will receive a set of pictures describing something that happened to a college student, Hiroshi. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Hiroshi to someone who has never seen the pictures.

You have 30 seconds to look through the pictures. There is no time limit to complete the task.

I encourage you to take time to think about the story while you are telling it. I will not remind you of doing so at this time, but please do try to take time to think about what you are going to say.

Once again, thank you for your cooperation!

G6: Familiarization task
Thank you for participating in this study. This is a practice session. You will receive a set of pictures describing something that happened to Yoshio. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Yoshio to someone who has never seen the pictures.

You have 5 minutes to prepare for this task. You may take notes on the sheet of paper provided, but please do not write out everything you intend to say either in English or in Japanese. The paper will be removed at the end of the preparation time and you will not have it while you are speaking. There is no time limit to complete the task.

I encourage you to think about the story while you are telling it. To remind you of doing so, I will tell you “don’t forget that you can take time to think about what you are going to say” during the task.

G6: Main task
Thank you for participating in this study. You will receive a set of pictures describing something that happened to a college student, Hiroshi. Please tell the story in Japanese in as much detail as possible as if you were describing what happened to Hiroshi to someone who has never seen the pictures.

You have 10 minutes to prepare for this task. You may take notes on the sheet of paper provided, but please do not write out everything you intend to say either in English or in Japanese. The paper will be removed at the end of the preparation time and you will not have it while you are speaking. There is no time limit to complete the task.

I encourage you to take time to think about the story while you are telling it. I will not remind you of doing so at this time, but please do try to think about what you are going to say.

Once again, thank you for your cooperation!
APPENDIX D: SAMPLE QUESTIONS IN THE
SEMISTRUCTURED RETROSPECTIVE INTERVIEW

1. Warm-up questions:
   How did you feel about the task?
   How did you feel about your story while you were telling it?

2. Questions regarding pre-task planning (Groups 3, 4, & 6 only):
   How did you spend your time?
   What did you do during the preparation time?
   What would you say your focus was while you were preparing your story?
   (Looking at the participant’s notes) Why did you write the outline/write out everything?

3. Questions regarding on-line planning:
   (Groups 2 & 4 only)
   How did you feel about the time limit?
   Would your story have been different if there had not been the time limit?
   (all the groups)
   Were you aware of thinking while you were telling your story?
   What were you thinking while you were telling the story?
   Did you spend time analyzing and revising parts of your story after you said it?

4. Questions focusing on the story line and prompts:
   What was your reaction when you first looked at the pictures?
   Did you have trouble telling the story?

5. Questions while watching learner behavior in the video:
   What were you doing here?
   What were you thinking at this time?
   (When the participant’s stopped talking in the video) What did you focus on at that time?

6. Reactive questions (following up the participant’s comment):
   What do you mean (by that)?
   Why?
   Can you explain that a little bit more?
   What was difficult about it?

7. General questions:
   How do you feel about speaking in Japanese in general?
   Anything else you would like to tell me on how you felt during the task?
APPENDIX E: STIMULATED RECALL INSTRUCTION

I am going to ask you questions while we watch the video. I am interested in what you were thinking at the time you were preparing for the task and also while you were talking about the pictures. We can see what you were doing, but I don’t know what you were thinking. So, what I’d like you to do is tell me what you were thinking, what was in your mind at that time, but not what you are thinking now.

I’m going to put the remote control on the table here and you can pause the video any time you want. So if you want to tell me something about what you were thinking, you can push pause. If I have a question about what you were thinking, then I will push pause and ask you to talk about that part of the video. You may answer, “I don’t remember,” if you are not sure what you were thinking at that time.
APPENDIX F: FACTOR ANALYSIS

<table>
<thead>
<tr>
<th>Metric</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
<th>Component 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>.202</td>
<td>.940</td>
<td>.032</td>
<td>-.125</td>
</tr>
<tr>
<td>Pruned speech rate</td>
<td>.178</td>
<td>.938</td>
<td>.056</td>
<td>-.109</td>
</tr>
<tr>
<td>Pauses</td>
<td>-.089</td>
<td>-.857</td>
<td>-.108</td>
<td>-.030</td>
</tr>
<tr>
<td>Clauses/T-unit</td>
<td>.848</td>
<td>.129</td>
<td>.114</td>
<td>.104</td>
</tr>
<tr>
<td>Words/T-unit</td>
<td>.861</td>
<td>.098</td>
<td>.097</td>
<td>-.148</td>
</tr>
<tr>
<td>TTR</td>
<td>.110</td>
<td>-.046</td>
<td>-.012</td>
<td>.869</td>
</tr>
<tr>
<td>Modified TTR</td>
<td>.744</td>
<td>.198</td>
<td>.050</td>
<td>.087</td>
</tr>
<tr>
<td>Error-free clauses</td>
<td>.546</td>
<td>.150</td>
<td>.546</td>
<td>-.272</td>
</tr>
<tr>
<td>Error rate</td>
<td>-.545</td>
<td>-.119</td>
<td>-.651</td>
<td>.400</td>
</tr>
<tr>
<td>Pruned error rate</td>
<td>-.494</td>
<td>-.146</td>
<td>-.641</td>
<td>.475</td>
</tr>
<tr>
<td>Correct particles</td>
<td>-.035</td>
<td>.032</td>
<td>.901</td>
<td>.109</td>
</tr>
<tr>
<td>Correct verb forms</td>
<td>.482</td>
<td>.173</td>
<td>.220</td>
<td>-.453</td>
</tr>
</tbody>
</table>
APPENDIX G: CORRELATIONS AMONG MEASURES
<table>
<thead>
<tr>
<th></th>
<th>Speech rate</th>
<th>Pruned speech rate</th>
<th>Pauses</th>
<th>Clauses/T</th>
<th>Words/T</th>
<th>TTR</th>
<th>Modified TTR</th>
<th>Error-free</th>
<th>Error rate</th>
<th>Pruned error rate</th>
<th>Correct particle</th>
<th>Correct verb form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech rate</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pruned speech rate</td>
<td>.973**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pauses</td>
<td>-.720**</td>
<td>-.703**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clauses/T</td>
<td>.276**</td>
<td>.261**</td>
<td>-.247**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words/T</td>
<td>.289**</td>
<td>.266**</td>
<td>-.205**</td>
<td>.855**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>-.115</td>
<td>-.105</td>
<td>.052</td>
<td>.028</td>
<td>-.147</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified TTR</td>
<td>.323**</td>
<td>.316**</td>
<td>-.195**</td>
<td>.496**</td>
<td>.485**</td>
<td>.108</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>.290**</td>
<td>.271**</td>
<td>-.260**</td>
<td>.446**</td>
<td>.489**</td>
<td>-.118</td>
<td>.426**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error rate</td>
<td>-.305**</td>
<td>-.305**</td>
<td>.178</td>
<td>-.431**</td>
<td>-.553**</td>
<td>.214</td>
<td>-.457**</td>
<td>-.750**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pruned error rate</td>
<td>-.325**</td>
<td>-.326**</td>
<td>.204</td>
<td>-.424**</td>
<td>-.540**</td>
<td>.321</td>
<td>-.405**</td>
<td>-.734**</td>
<td>.965**</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct particle</td>
<td>.045</td>
<td>.073</td>
<td>-.117</td>
<td>.203*</td>
<td>.152</td>
<td>-.077</td>
<td>.039</td>
<td>.325**</td>
<td>-.438**</td>
<td>-.439**</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Correct verb form</td>
<td>.308**</td>
<td>.292**</td>
<td>-.201*</td>
<td>.294**</td>
<td>.398**</td>
<td>-.139</td>
<td>.341**</td>
<td>.544**</td>
<td>-.598**</td>
<td>-.556**</td>
<td>.098</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
REFERENCES


