

<http://researchspace.auckland.ac.nz>

*ResearchSpace@Auckland*

### **Copyright Statement**

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.

<http://researchspace.auckland.ac.nz/feedback>

### **General copyright and disclaimer**

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

# THE EFFECTS OF TASK PLANNING ON L2 WRITING

LIN, Yan

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF DOCTOR OF PHILOSOPHY IN LANGUAGE TEACHING AND  
LEARNING

THE UNIVERSITY OF AUCKLAND, 2013

DEPARTMENT OF APPLIED LANGUAGE STUDIES AND LINGUISTICS



## ABSTRACT

This thesis reports a study that investigated the effects of pre-task planning and online planning on L2 writing performance and L2 writing development. Learners' pre-task planning and while-writing processes were also examined. It aimed to produce findings that would contribute to the theoretical issues of how allocation of attentional resources affects learners' language use and how practicing in different task conditions influences L2 development as well as pedagogical issues of how tasks can be manipulated in L2 classrooms to promote L2 learning. It attempted to fill the research gaps in task planning literature in terms of the longer-term effects of planning, which has been lacking, the effects of planning on written production, which is scanty, and how pre-task planning affects during-task attentional allocation.

The study was conducted in a Chinese university and seventy-five non-English major freshmen participated in the study. At the beginning of the study, the participants all completed a writing task in 25 minutes, which served as a pre-test. They were then assigned into three groups for treatment: a no planning group (NP), which had 15 minutes to complete each writing task; an online planning group (OLP), which had 25 minutes to complete each task, and a pre-task planning group (PTP), which had 10 minutes to plan before completing the task in 15 minutes. They completed four experimental writing tasks under different task conditions on a weekly basis. The four experimental tasks were all compare/contrast type of essays but with different topics. A week after the last experimental task, participants wrote another writing task all under the same writing condition, which served as a posttest. In the following week, a post-task questionnaire survey and an interview were administered to collect data on participants' pre-task planning and while-writing processes. Ten weeks later, a delayed posttest was administered in the same way as the posttest. Learners' written production was measured in terms of accuracy, complexity, fluency, and organization.

Results revealed that neither pre-task planning nor online planning affected L2 writing performance in terms of accuracy, syntactic complexity, or organization of the

essays. Pre-task planning had a positive effect on lexical complexity. With respect to fluency, pre-task planning had a positive effect on repair fluency (i.e. dysfluency) while online planning had a negative effect on temporal fluency (i.e. writing speed). With regard to the effects of planning on L2 writing development, both pre-task planning and online planning had some beneficial effects on accuracy over time but no effect on complexity. Both pre-task planning and online planning had negative effects on temporal fluency and no effect on organization. The analyses of questionnaire and interview data indicated that pre-task planning provided more space for during-task attention to form and the online monitoring and editing may raise learners' awareness of certain problematic linguistic form(s).

## ACKNOWLEDGEMENTS

This thesis would not have been possible without the help of many individuals. In particular I would like to thank the following.

First and foremost, my two supervisors—Dr. Rosemary Erlam and Dr. Rob Batstone, for the guidance and valuable suggestions they have provided on the study design and the insightful comments on the dissertation.

Professor Rod Ellis and Professor Peter Skehan for their valuable suggestions on study design and data analysis.

Xiaohong Wang and Wenjing Lin for allowing me to conduct the study in their classes and assisting with data collection. My gratitude also goes to the students who participated in this study. Without their support this study would not have been possible.

The deans of the School of English and International Studies, Beijing Foreign Studies University, who have given me great support during my four-year study.

My colleagues in Beijing Foreign Studies University, Professor Yi'an Wu, Luxin Yang, Jiajin Xu, Baocheng Han, Maocheng Liang, to name just a few, who have given me valuable advice at different stages of the study.

My friends and fellow PhD students in the Department of Language Studies and Linguistics of the University of Auckland, who have given me advice and all kinds of help. Their friendship and support have made my study and stay in New Zealand enjoyable.

Kay Irwin, Xiaoying Wang, Jinzhao Li, and Dai Ning, who helped with data coding. Brian McArdle for advising me on statistical analysis and Russell Greenwood, who proofread my thesis.

Last but not the least, my family, whose unconditional love and support helped me get this far.



## TABLE OF CONTENTS

ABSTRACT .....	iii
ACKNOWLEDGEMENTS .....	v
TABLE OF CONTENTS .....	vii
LIST OF TABLES .....	xv
LIST OF FIGURES .....	xviii
CHAPTER 1 INTRODUCTION .....	1
1.1 Overview .....	1
1.2 Motivation for the present study .....	3
1.3 Purpose and significance of the study .....	4
1.4 Context of the study .....	6
1.5 Thesis outline .....	7
CHAPTER 2 THEORETICAL BACKGROUND .....	9
2.1 Overview .....	9
2.2 Information processing framework .....	10
2.2.1 Processing limitations .....	11
2.2.2 Attention .....	11
2.2.3 Memory system .....	12
2.2.4 Controlled processing and automatic processing .....	14
2.2.5 Anderson's ACT* .....	16
2.3 Information processing and language learning .....	18
2.3.1 Information-processing perspective of second language learning .....	18
2.3.2 Input .....	21
2.3.3 Central processing .....	22
2.3.4 Output .....	23
2.3.4.1 The Output Hypothesis .....	23
2.3.4.2 Speech production model .....	25
2.3.4.3 Models of writing .....	27



2.4 Task-based language teaching and learning.....	32
2.4.1 Definition of ‘task’ .....	34
2.4.2 Task complexity .....	35
2.4.3 Task implementation .....	38
2.5 Summary .....	39
<b>CHAPTER 3 THE EFFECTS OF PLANNING ON LANGUAGE PRODUCTION: A LITERATURE REVIEW .....</b>	<b>43</b>
3.1 Overview.....	43
3.2 The nature of planning .....	43
3.3 Classification of planning .....	45
3.3.1 Pre-task planning .....	46
3.3.2 Within-task planning.....	47
3.4 Measures for task-based language performance .....	48
3.5 The effects of task planning on oral production .....	49
3.5.1 The effects of rehearsal on oral production .....	49
3.5.2 The effects of strategic planning on oral production: introduction	52
3.5.3 The effects of strategic planning on oral production: accuracy .....	52
3.5.4 The effects of strategic planning on oral production: complexity	54
3.5.5 The effects of strategic planning on oral production: fluency .....	55
3.5.6 The effects of online planning on oral production.....	56
3.5.7 The process of planning: how do learners plan .....	57
3.5.8 Summary.....	58
3.6 Impact of planning on L1 writing .....	59
3.7 Comparison of planning in L1 and L2 writing .....	64
3.8 Impact of planning on L2 writing .....	64
3.8.1 Impact of pre-task planning on L2 writing: accuracy.....	65
3.8.2 Impact of pre-task planning on L2 writing: complexity .....	66
3.8.3 Impact of pre-task planning on L2 writing: fluency .....	67
3.8.4 Impact of online planning on L2 writing .....	68

3.8.5 Summary .....	78
3.9 Summary of review .....	79
3.10 Research gaps the present study addresses .....	80
CHAPTER 4 METHODOLOGY .....	81
4.1 Overview .....	81
4.2 Research questions .....	81
4.3 Overview of the experimental design .....	82
4.4 Participants.....	83
4.4.1 Selection of participants and their proficiency level.....	83
4.4.2 Demographic characteristics of participants .....	85
4.5 Operationalization of planning conditions.....	85
4.5.1 Operationalization of pre-task planning.....	85
4.5.2 Operationalization of online planning.....	87
4.6 Instruments.....	88
4.6.1 The writing tasks .....	88
4.6.1.1 The target structures .....	89
4.6.1.2 Task design.....	91
4.6.1.3 Description of the tasks.....	92
4.6.1.4 Task complexity .....	93
4.6.2 Questionnaires.....	96
4.6.3 Semi-structured Interview.....	97
4.7 Procedures for data collection.....	98
4.7.1 Selection of research site and participants .....	98
4.7.2 Assigning participants into research conditions.....	99
4.7.3 Procedures for the testing sessions.....	100
4.7.4 Experimental procedures.....	100
4.7.4.1 Procedures for the pre-task planning group.....	101
4.7.4.2 Procedures for the online planning group .....	102
4.7.4.3 Procedures for the no planning group .....	102

4.7.5 Procedures for the post-task questionnaire and the post-study questionnaire .....	102
4.7.6 Procedures for the post-task interview .....	103
4.8 Data coding and scoring procedures .....	103
4.8.1 The writing tasks .....	103
4.8.1.1 The accuracy measures .....	104
4.8.1.2 The complexity measures .....	105
4.8.1.3 Measures for fluency .....	110
4.8.1.4 Measures for the organization of the essays .....	110
4.8.2 Questionnaires and interviews .....	116
4.8.3 Planning notes .....	118
4.9 Data Analysis .....	119
4.9.1 Analysis of learners' writing performance .....	119
4.9.2 Analysis of planning notes .....	120
4.9.3 Analysis of questionnaire and interview data .....	121
4.9.4 Inter-rater reliability analysis .....	121
4. 10 Pilot study .....	123

## CHAPTER 5 THE EFFECTS OF TASK PLANNING ON IMMEDIATE L2 WRITING

PERFORMANCE .....	129
5.1 Overview .....	129
5.2 Overview of data analysis .....	129
5.3 Results for the independent variable .....	132
5.4 Results for the dependent variables .....	132
5.4.1 Results for accuracy measures .....	133
5.4.2 Results for syntactic complexity measures .....	135
5.4.3 Results for lexical complexity measures .....	136
5.4.4 Results for fluency measures .....	138
5.4.5 Results for organization scores .....	140
5.5 Summary .....	142

5.6 Discussion .....	143
5.6.1 Accuracy.....	144
5.6.1.1 Effects of pre-task planning on accuracy.....	144
5.6.1.2 The effects of online planning on accuracy .....	149
5.6.2 Complexity.....	152
5.6.2.1 Effects of pre-task planning on complexity .....	152
5.6.2.2 Effects of online planning on complexity .....	159
5.6.3 Fluency.....	161
5.6.3.1 Effects of pre-task planning on fluency .....	161
5.6.3.2 Effects of online planning on fluency .....	164
5.6.4 Organization.....	165
5.6.4.1 Effects of pre-task planning on organization .....	165
5.6.4.2 Effects of online planning on organization .....	167
5.6.6 Summary .....	167
 CHAPTER 6 THE EFFECTS OF TASK PLANNING ON L2 WRITING DEVELOPMENT .....	 171
6.1 Overview .....	171
6.2 Results of the pre-test.....	171
6.3 Changes over time.....	173
6.3.1 Changes in accuracy over time .....	173
6.3.2 Changes in syntactic complexity over time .....	175
6.3.3 Changes in lexical complexity over time.....	177
6.3.4 Changes in fluency over time.....	177
6.3.5 Changes in organization scores over time.....	179
6.3.6 Summary .....	180
6.4 Between-group differences .....	181
6.4.1 Between-group differences in accuracy .....	182
6.4.2 Between-group differences in syntactic complexity.....	183
6.4.3 Between-group differences in lexical complexity.....	183

6.4.4	Between-group differences in fluency .....	184
6.4.5	Between-group differences in organization .....	185
6.5	Summary of results .....	186
6.6	Discussion.....	188
6.6.1	Longer-term effects of pre-task planning.....	188
6.6.2	Longer-term effects of online planning .....	195
6.6.3	Summary.....	198
CHAPTER 7 THE PROCESS OF PLANNING AND WRITING.....		201
7.1	Overview.....	201
7.2	Overview of data analysis.....	201
7.3	The process of pre-task planning .....	202
7.3.1	Focus of planning.....	202
7.3.2	Planning strategies .....	205
7.4	Plan application and planning styles.....	211
7.4.1	Plan application.....	211
7.4.2	Planning styles .....	214
7.5	The process of writing for the tasks .....	215
7.5.1	Allocation of writing time.....	215
7.5.2	While-writing focus .....	216
7.5.3	Strategies employed while writing .....	220
7.6	Learners' comments on task completion conditions.....	224
7.6.1	Learners' perceptions of pre-task planning.....	224
7.6.2	Learners' perception on the online planning and no planning conditions .....	226
7.7	Summary of results .....	228
7.8	Discussion.....	229
7.8.1	Operationalization of planning .....	229
7.8.2	The process of planning.....	231
7.8.3	Students' perceived benefits of pre-task planning .....	234

7.8.4 The process of writing.....	235
7.8.5 The influence of the social context .....	239
7.8.5.1 Chinese students' lack of training on planning .....	239
7.8.5.2 The influence of Chinese educational culture.....	240
CHAPTER 8 CONCLUSION.....	245
8.1 Overview .....	245
8.2 Summary of main findings.....	245
8.2.1 The immediate effects of planning.....	246
8.2.2 The effects of planning on L2 writing development.....	252
8.2.3 What learners do during planning and writing.....	254
8.3 Theoretical implications.....	255
8.3.1 L2 skill development.....	256
8.3.2 Attention.....	257
8.3.3 The Output Hypothesis .....	258
8.3.4 Task complexity .....	259
8.4 Methodological issues raised by the study.....	260
8.4.1 Difficulty of operationalization of planning .....	260
8.4.2 Difficulty of measuring L2 writing performance.....	260
8.4.3 Difficulty of conducting semi-longitudinal quasi-experimental research .....	262
8.4.4 Use of focused tasks.....	262
8.4.5 Influence of sociocultural factors.....	263
8.5 Pedagogical implications .....	263
8.6 Limitations of the present study.....	265
8.7 Suggestions for future research.....	267
LIST OF REFERENCES .....	269
APPENDIX A: CONSENT FORM .....	284
APPENDIX B: QUESTIONNAIRE FOR BACKGROUND INFORMATION .	285
APPENDIX C: WRITING TASKS FOR THE TESTING SESSIONS .....	286

APPENDIX D: WRITING TASKS FOR THE TREATMENT SESSIONS.....	289
APPENDIX E: POST-TASK QUESTIONNAIRE .....	292
APPENDIX F: POST-TASK INTERVIEW QUESTIONS .....	297
APPENDIX G: POST-STUDY QUESTIONNAIRE .....	299
APPENDIX H: COH-METRIX INDEXES .....	300
APPENDIX I: SLA STUDIES ON PRE-TASK PLANNING USING ORAL PRODUCTION DATA .....	302

## LIST OF TABLES

Table 1 Studies that investigated the effects of planning on L2 written production....	71
Table 2 Overview of the experimental design.....	83
Table 3 Demographic information of the participants .....	85
Table 4 Features of the tasks in the present study .....	96
Table 5 Summary of Coh-Metrix variables included in the present study.....	115
Table 6 Examples of four types of plan .....	119
Table 7 Summary of dependent variables .....	130
Table 8 Interpretation of effect sizes .....	132
Table 9 Results for the total number of words produced in the experimental sessions	132
Table 10 Descriptive statistics for the accuracy measures in the treatment sessions.	133
Table 11 Results for the target structures in the treatment sessions.....	134
Table 12 Results for the general measures for accuracy in the treatment sessions.....	134
Table 13 Descriptive statistics for syntactic complexity in the treatment sessions....	135
Table 14 Results for syntactic complexity measures in the treatment sessions .....	135
Table 15 Descriptive statistics for lexical complexity in the experimental sessions .	136
Table 16 Results for lexical complexity in the treatment sessions .....	137
Table 17 Pairwise comparison results of lexical sophistication (Lam) in the treatment sessions.....	137
Table 18 Descriptive statistics for fluency measures in the treatment sessions.....	138
Table 19 Results for fluency measures in the treatment sessions .....	138
Table 20 Results of post hoc comparisons of WMP in the treatment sessions .....	139
Table 21 Repeated measures ANOVA results of time effect on WPM by group.....	139
Table 22 Results of post hoc comparisons of Dysfluency in the treatment sessions.	139
Table 23 Descriptive statistics for organization measures in the treatment sessions.	140
Table 24 Results for local and global coherence and cohesion in the treatment sessions.....	141
Table 25 Results for subjective ratings on organization in the treatment sessions ....	141



Table 26 Mann-Whitney results of subjective ratings in Time 4 writing .....	141
Table 27 Wilcoxon signed ranks results of subjective ratings in treatment sessions.	142
Table 28 Summary of significant between-group and within-group differences .....	143
Table 29 ANOVA results of the general measures in the pre-test.....	172
Table 30 Descriptive statistics for the accuracy measures in the three tests.....	173
Table 31 Results for accuracy measures across tests .....	174
Table 32 Descriptive statistics for syntactic complexity in the three tests .....	175
Table 33 Results for syntactic complexity measures across tests .....	176
Table 34 Descriptive statistics for lexical complexity in the three tests .....	177
Table 35 Results for lexical complexity across tests .....	177
Table 36 Descriptive statistics for fluency in the three tests.....	178
Table 37 Results for fluency measures across tests .....	178
Table 38 Descriptive statistics for organization in the tests.....	180
Table 39 Results for organization measures across tests .....	180
Table 40 Between-group differences in accuracy in the posttests .....	182
Table 41 Between-group differences in syntactic complexity in the posttests .....	183
Table 42 Between-group differences in lexical complexity in the posttests.....	184
Table 43 Between-group differences in fluency in the posttests .....	185
Table 44 Between-group differences in organization in the posttests .....	186
Table 45 Summary of significant between-group and within group differences in the posttests.....	187
Table 46 Summary of data used to answer Research Question 3 .....	202
Table 47 Questionnaire results for focus of planning by pre-task planners.....	204
Table 48 Summary of planning strategies reported in the present study .....	207
Table 49 Planning strategies reported in the questionnaire and the interview.....	209
Table 50 Application of plans reported in the questionnaire .....	212
Table 51 Descriptive statistics of the total number of words on the planning sheets and the percentage of the planned words applied in the essays.....	213
Table 52 Repeated measures ANOVA results for total number of words on planning	

notes and percentage of planned words applied in the essays .....	213
Table 53 Raw number of planning styles used in the treatment sessions .....	214
Table 54 While-writing focus reported by participants.....	217
Table 55 Reported strategies used during writing.....	221
Table 56 Summary of benefits of pre-task planning identified by learners .....	225
Table 57 Reported benefits of pre-task planning .....	225

## **LIST OF FIGURES**

Figure 1 Kellogg's model of working memory in writing.....	30
Figure 2 Stages in task implementation (modified based on Skehan, 1996a, 1996b) .	38
Figure 3 Overview of the pilot study .....	124

# CHAPTER 1 INTRODUCTION

## 1.1 Overview

The aim of the present study is to explore the effects of planning on second language (L2) writing performance and L2 writing development. In the literature of planning studies, ‘planning’ has been defined as “essentially a problem solving activity” (R. Ellis, 2005b, p. 3), which involves a wide range of mental activities, such as identifying problems, setting goals, reflecting on past experience, and devising strategies (Wendel, 1997). In language production, it refers mainly to devising strategies for accomplishing a task or a communicative act, conceptualizing the propositional content, and retrieving and formulating linguistic forms that are necessary for meaning expression.

Planning has attracted considerable attention in second language acquisition (SLA) research and has become an area of inquiry in its own right (Ortega, 2005). There are two principal reasons for the interest in planning in SLA research: theoretical and pedagogical. Theoretically, planning connects with the constructs of processing capacity, attention, focus on form, declarative/procedural knowledge, and controlled processing and automatic processing, which will be discussed in the next chapter. It offers an area in which the role of attention in SLA can be studied. When learners engage in strategic planning before a task or employ careful within-task planning, opportunities to attend to language arise. This may affect the way learners use language. Furthermore, it is hypothesized that the kind of language learners employ in production will influence the process of interlanguage (IL) development (e.g. R. Ellis 2005b). According to research that has attempted to address the relation between task planning and IL development (e.g. Crookes, 1989; R. Ellis, 1987; Skehan, 1998), there is a facilitative role for planning in IL development. Crookes (1989), argued from an information processing perspective that under a planning condition learners could push the interlanguage to its limits and thus engage second language learning processes. That is, planning enables learners to

manipulate their attentional resources so that they can put to use linguistic forms which are on the edge of their current capacity. R. Ellis (1987), from a variationist point of view, proposed that planning should create an opportunity for the not yet fully automatized linguistic forms to be accessed and internalized so that they can later be incorporated into the vernacular style<sup>1</sup> (Tarone, 1979, 1983, 1985). If task planning is used systematically in a classroom, they should enable learners to become more comfortable using linguistic forms which are at the upper limit of their current capacity, allowing them to become more automatized and more easily accessed in real-time communication. In the present study, interlanguage is defined as a learner's L2 system that is manifested by his/her "linguistic knowledge and the control of that knowledge" (Bialystok & Smith, 1985, p. 101) in language use. Second language learning is seen as the process of developing the IL system by gaining more linguistic knowledge and/or more control of that knowledge. Development in L2 writing indicates and demonstrates changes in IL.

Apart from the theoretical motivation for studies on planning in SLA, there is also a pedagogical motivation. This is because planning is a procedure that can be implemented in language teaching. It is known that communication in L2, be it oral or written, may pose difficulty on the part of L2 learners because of their limited linguistic resources. This is particularly the case when L2 learners perform tasks that are cognitively demanding under time pressure. To help learners deal with this difficulty, teachers can provide learners with planning time before and/or during the task. Research has shown that planning can help reduce the cognitive load of a task and alleviate time pressure, thus may bring about a more successful language performance in terms of complexity, accuracy, and/or fluency (see Chapter 3). Therefore, planning is a device that language teachers can use to influence the process of language production, in L1 as well as in L2, for the purpose of promoting learning. However, planning needs to be implemented systematically. How it is implemented may affect its effectiveness. Results from studies

---

<sup>1</sup>According to Tarone (1979, 1983), interlanguage consists of a continuum of styles which is defined by the degree of attention paid to speech. At one end is the careful style or attended style and at the other is the vernacular style, where the least amount of attention is given to speech.

on planning can provide information that helps language teachers make pedagogical decisions, such as whether or not to provide learners with the opportunity to plan before production, when to provide the opportunity, and how to implement it.

## **1.2 Motivation for the present study**

In general, the present study was conducted out of an interest in the theoretical and pedagogical issues that motivated other planning studies in SLA. More specifically, a review of the existing literature on planning (see Chapter 3 for more detailed review) found some gaps that need to be filled. The first gap is that SLA planning studies have focused more on the effects of planning on oral production, rendering planning effects on written production unclear. While results for the effects of planning on oral production shed light on how the mental activities learners engage in during the process of production influence the outcome, the picture will not be complete until more is known about how planning affects written production.

The second gap is that in the existing literature there is little empirical evidence that can directly support the argument for a positive role for planning in IL development. Though there has been some evidence to suggest a facilitative role of planning in language learning (Crookes, 1989; R. Ellis, 1987), SLA studies on planning have only investigated the immediate effects of planning on task performance to date. As a language teacher, the researcher has often asked students to produce planned output in speaking and writing classes, thinking that producing planned output could facilitate IL development. However, there has not been empirical evidence that enables the researcher to provide an answer to the question imposed by students---“Can my English be really improved through practicing prepared speech and writing?” with confidence. As an attempt to answer this question, this study goes some way towards investigating the longer-term effects of planning by using a delayed posttest of learners’ writing ability.

### 1.3 Purpose and significance of the study

The main objectives of the study are: 1) to examine the effects of task planning on immediate L2 writing performance; 2) to investigate whether task planning leads to any change in L2 writing over time; and 3) to investigate what learners do during pre-task planning time and how learners under different task conditions go about completing the writing tasks. The first objective was met by comparing participants' written texts produced under planning conditions and those under a no planning condition in the treatment sessions. The second objective was achieved by examining whether there were changes in participants' writing in the posttests compared with the pre-test (see Chapter 4 for detailed description). The third objective was achieved through analysis of participants' self-reports.

The significance of the study lies mainly in two aspects: contribution to the existing literature on task planning and pedagogical implications. First, the study contributes to the existing literature in the following ways:

1) This study presents evidence that has implications for the connection between task planning and IL development. As has been pointed out earlier, assumptions have been made in previous planning studies about a facilitative role of planning in IL development based on comparisons of planned output with unplanned output (e.g. Crookes, 1989; R. Ellis, 1987). The designs of these studies, however, did not allow for an examination of whether the immediate effects of planning (i.e. increased accuracy, complexity, or fluency) could be carried over to a new task or to a later time. No solid evidence of the beneficial effects of task planning on IL development has been obtained. Therefore, the claim that task planning facilitates language development still remains a hypothesis. The present study overcomes this limitation, to some degree, by adopting a pretest-posttest-delayed posttest design so that learners' performance in a new context (i.e. a new task and a new task condition) can be examined. Moreover, the present study extended over a longer time period (i.e. 14 weeks) than previous studies, so whether planning causes any change in L2 writing over time could be examined.

2) The present study investigates the effects of planning on written production, which is an under-researched area. In the existing literature on task planning, most studies have investigated the effects of planning on oral production. The number of studies that looked at the effects of planning on written production is much smaller than that on oral production, which will become evident in Chapter 3 when previous studies are reviewed. The small number of studies makes it difficult to draw conclusions on how planning affects L2 writing. This warrants more studies to be conducted investigating the effects of planning on L2 written production.

3) The present study looks at the effects of both pre-task planning and online planning. While the effects of pre-task planning have received much attention in previous planning studies, online planning is a relatively under-researched area (see Chapter 3 for a review of these studies). A few studies (e.g. D. Li, 2004; Nakakubo, 2011; Wang, 2009; Yuan, 2001) used online planning as an independent variable and these studies produced different results for the effects of online planning. This necessitates further research on online planning.

4) The present study not only investigates the process of pre-task planning but also the process of writing under different task conditions, which is an extension of previous studies on the process of planning. Although studies on planning have been growing in number, investigation into the process of planning is curiously still scarce. Only a few studies (e.g. Guará-Tavares, 2008; Nakakubo, 2011; Ortega, 1999; Sangarun, 2001) investigated how learners responded to the planning opportunity. These studies only investigated what learners do during pre-task planning time leaving how pre-task planning influenced the subsequent production process and how the pre-task planners' processes of production might be different from those who carry out language tasks under a no planning condition unexplored. The present study extends the previous studies by investigating not only what learners do during pre-task planning but also what they do while they are completing the writing tasks.

Theoretically, the present study aims to further our understanding of how learners



allocate their attentional resources under different task conditions in L2 production; and how this allocation of attention in turn affects language use and changes in IL. In this way, the present study may contribute to theorizing L2 acquisition in terms of information processing theory (R. Ellis, 2005b).

The present study also has implications for L2 pedagogy because of the reasons stated earlier in Section 1.1. That is, planning could be a pedagogical device that is implemented in the classroom to influence the kind of language learners produce. It is a tool that can be employed by language teachers to engage learners in processes that may lead to interlanguage development. More specifically, results of the present study may provide information on what kind of writing performance could be expected when L2 learners are given time to plan. Based on this information, language teachers could make decisions on when to provide planning opportunities, depending on their teaching objectives. Results for learners' processes of planning and writing may offer information that could help teachers decide whether or not guidance or training on planning should be provided and how to do it.

#### **1.4 Context of the study**

The study was carried out in China, where English is taught as the most important foreign language in schools and universities. The participants of the study were first-year university students studying in an English-as-a-foreign-language (EFL) context and they were at the intermediate level of English (see Chapter 4 Section 4.4 for more details on the participants). It is acknowledged that the particular characteristics of this group of participants, their English learning experiences, and individual differences may affect the results of the study. Hence, the conclusions should not be generalized to other contexts.

Conducting a study on planning in China is important. For one thing, China has a large population of English learners and English is one of the major subjects being taught in schools at various levels. The pedagogical implications of the present study may benefit a large number of people. For another, students' knowledge of how to plan

effectively is limited. Though they are often given some time to plan when they write in Chinese or English, how to make effective use of the time is not systematically taught. Moreover, a large amount of writing practice, particularly in high schools, is for the purpose of passing exams, in which students would not have much time to plan. If the present study can find out how planning influences L2 performance and L2 learning, it will have importance implications for English teaching in China.

### **1.5 Thesis outline**

This dissertation is organized into eight chapters. Chapter 2 lays the theoretical ground for the present study and Chapter 3 reviews studies that have been previously conducted on task planning. Chapter 4 describes the method for data collection and data analysis and presents the research questions and hypotheses. Chapter 5 answers the question regarding the effects of planning on immediate L2 writing performance and Chapter 6 addresses the question regarding the effects of planning on L2 writing development. Chapter 7 presents the results of qualitative analysis of the questionnaire and interview data that reveals what learners do before and during task completion. Results from an analysis of learners' planning sheets will also be presented. Chapter 8 produces the conclusions, implications, and limitations of the study.



## CHAPTER 2 THEORETICAL BACKGROUND

### 2.1 Overview

This chapter outlines the theoretical and pedagogical background of the study. As explained in Chapter 1, planning engages cognitive processing in language use, therefore the study of task planning in second language acquisition research has largely drawn on information processing models of learning. In information processing theories, humans are considered to have limited processing ability and attentional capacity. Therefore, language users, particularly L2 learners, cannot attend to meaning and form at the same time to the same degree. Planning is believed to have the potential of mitigating these limitations by allowing L2 learners to attend to one or both of these aspects of communication (i.e. meaning and form) before production so more attention to language form is possible during production, thus leading to improved language use. Moreover, because of the potential positive effects on language production, it is hypothesized that planning is facilitative to L2 development (e.g. Crookes, 1989, R. Ellis, 1987). From the discussion above it can be seen that planning is connected with the following theoretical constructs involved in information processing models: processing limitations, attention, memory, and controlled processing and automatic processing, which will be discussed in more detail in the following sections of this chapter. Apart from the above theoretical background, this chapter also discusses the pedagogical background of the present study ---task-based language teaching and learning.

This chapter begins with a very brief discussion of the information processing account of learning, with a focus on the theoretical constructs outlined above. Section 2.3 relates information processing to language learning by discussing how the stages of information processing (i.e. input, central processing, and output) have been applied to explain the process of learning a language. The output stage will be the main focus of discussion in this section since the present study involves research relating to the production of language. In section 2.4, task-based instruction, which provides the

pedagogical context for the present study, will be discussed.

## **2.2 Information processing framework**

The idea that planning can be useful in language learning has its origin in research undertaken based on an information processing framework, which essentially explores how attention, a limited but also a vital resource for engaging with language, can be supported in various ways so as to allow for a more focal attention to language. One of these ways is through planning. Given the centrality of information processing perspectives it is important to review their main features.

Information processing is a general term for a framework used to explain how the human mind converts information it picks up from the environment for use. In other words, information processing theory, taking advantage of using computer and computational processes to model the human brain and mind respectively, provides an account of how information is stored and transformed in the human mind and how the brain retrieves information for output. According to this theory (e.g. Atkinson & Shiffrin, 1968; Craik & Lockhart, 1972; Rumelhart & McClelland, 1986), complex human behavior builds on simple steps that can be isolated and studied. In general, three basic stages, which interact with one another, are identified in information processing. They are input, central processing, and output. Input refers to the input material available in the environment, which consists of things that are perceivable by our senses. The input that we pay attention to is picked up and processed through central processing mechanisms (i.e. memory system). After being processed, it goes to the output stage.

Though there are different models of information processing (e.g. Atkinson & Shiffrin, 1968; Craik & Lockhart, 1972; Rumelhart & McClelland, 1986), some basic principles can be summarized. The first is that the human organism is genetically predisposed to process and organize information in specific ways. The second principle is that human beings are limited in their processing capacity. The third is that mental processing comprises two different types of processes, automatic and controlled (e.g.

Shiffrin & Schneider, 1977). Another principle of information processing is that there is a control mechanism named the central executive<sup>2</sup>(Baddley, 1986), which supervises cognitive processes, such as planning, rule acquisition, selection, initiation and termination of process routines, etc. The above-mentioned principles and the key constructs involved in the discussion of these assumptions will be laid out in the following sections.

### **2.2.1 Processing limitations**

In an information processing framework, humans are thought to be capacity-limited processors (Anderson, 1983; Newell & Simon, 1972). This limitation can be described according to two dimensions: processing ability and focus of attention. For the first dimension, humans are thought to be limited in their ability to process information. An individual is able to process information to a greater or lesser extent based on his/her past experience and expectancies. As for the second dimension, it has been contended that human beings are not able to attend to all the input they are exposed to in the environment and have to direct their attention selectively to some of it at a given time. Attending to one thing means withdrawal of attention from some other things (W. James, 1890) in order to effectively process information. Since attention is of crucial importance in information processing, it will be discussed in more detail in the next section.

### **2.2.2 Attention**

According to James (1890), attention refers to “the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought” (p.403). How attention functions in the input stage can be explained in the following way. Under normal conditions, information can be sustained for only a few seconds in working memory (this will be explained further in the following section). However, if one piece of information comes to attention, it can be held in working

---

<sup>2</sup>It is a component of memory.

memory for longer, thus making it available for further processing and for entering into long-term memory (Shiffrin & Schneider, 1977). As has been pointed out in the last section, attention is capacity limited and selective to input from the environment (Broadbent, 1958). That is, we can only attend to a limited amount of material at a given time and information with sensory saliency (e.g. ambulance siren) is more likely to capture our attention, when attention is not under conscious control. When attention is under conscious control driven by goals or intentions of an individual, it is selective (W. James, 1890). Conscious decisions on what to pay attention to and the amount of resources to invest are based on demands of the task to be performed (Wickens, 1989). When we have to do concurrently two or more tasks that share similarities in terms of stages of processing or types (e.g. spatial or verbal, visual or auditory), performance on tasks deteriorates (ibid).

It needs to be noted that attention is not only important to input but also plays an important role at the output stage since it is a mechanism that controls the flow of information from input to output. It is essential for learning (Schmidt, 2001). The understanding of attention has informed research in second language acquisition (SLA), which will be elaborated in Section 2.3.3. Planning studies, in particular, have drawn on the concept of attention and become an area where the role of attention in SLA is studied.

For now the discussion will turn to memory, something that underlies all the principles of information processing.

### **2.2.3 Memory system**

In an account of the memory system, one theory that has received a lot of attention is the “modal model” (also known as “stage theory”) proposed by Atkinson and Shiffrin (1968). According to this theory, information is processed and stored in three stages: *sensory register*, *short-term store* (STS), and *long-term store* (LTS) (ibid). The environment makes available a lot of information (e.g. light, sound, smell, etc.) and the sensory information first enters the *sensory register*, resides there for a very brief period

of time, then decays and is lost. The information that receives attention is then transferred to the *short-term store*, which is the capacity for holding a small amount of information in an active and readily available state for a short period of time. The *short-term store* can hold information for 15 to 30 seconds (ibid), a period longer than it can be held in the sensory register, if the information is not repeated. Another feature of STS is that the number of units that can be processed at a time is limited. Miller (1956) gave the number as 7 (+/- 2) but more recent research suggests the number is more likely to be 5 (+/- 2) (Cowan, 2001). Therefore, the short-term store is capacity limited and temporary. In contrast, the *long-term store* holds the information transferred from the short-term store permanently. Its capacity is unlimited and long-lasting. This model proposes that information is processed in a serial manner as it moves from one stage to the next. That is, information passes from the sensory register to the short-term store and part of it may later be transferred to the long-term store.

In 1974, Baddeley and Hitch proposed a Model of Working Memory (Baddeley & Hitch, 1974), which suggested that the short-term store proposed by Atkinson and Shiffrin (1968) be replaced by a three-component working memory. There are three components in working memory: the central executive, the phonological loop, and the visuo-spatial sketchpad. The central executive is the supervisory system that controls information flow from and to its two slave systems (the phonological loop and the visuo-spatial sketchpad), which store short-term content-specific information (verbal and visuo-spatial). In 2000 Baddeley (Baddeley, 2000) added a third slave system to this model: the episodic buffer, which is responsible for linking phonological, visual and spatial information from the other two slave systems and binding them into a unitary episodic representation. The episodic buffer is also a workspace that allows access for long-term memory both for learning and retrieval (Baddeley, 2007).

Like the short-term store, the working memory capacity of a person is considered to be limited in terms of the amount of information and the length of time it can hold and process information. It is also noteworthy that research has shown that working memory



has a close link with attention (Desimone & Duncan, 1995), which has been discussed in Section 2.2.2. To reiterate, when not under conscious control our attention would be captured by information with sensory saliency. However, we can consciously shift our attention to input that needs to be processed in order to achieve our goal or intention. The goal-driven attention can override sensory-captured attention. The ability to override sensory-captured attention differs greatly between individuals and this difference is closely related to their working memory capacities (ibid). The greater a person's working memory the stronger the ability to resist sensory capture.

Previous studies, as will be seen in the next chapter, have provided some evidence to show that the limitation of working memory could be overcome by planning opportunities, which in turn affects language production. Now the discussion will turn to another pair of constructs: controlled processing and automatic processing.

#### **2.2.4 Controlled processing and automatic processing**

In discussion of human information processing, two different processing mechanisms, which are closely related to working memory and long-term memory, are differentiated: automatic processing and controlled processing (Shiffrin & Schneider, 1977). Human cognition is supported by both mechanisms and results from the interaction of the two kinds of processing. In doing a task, whether automatic processing or controlled processing is engaged depends on what type of memory (long-term memory or working memory) is utilized. An automatic process is defined as a sequence of relatively permanent set of associative nodes in long-term memory nodes that are always active in response to input (ibid). It is activated without requiring control or attention of a person. Since little attention is needed in automatic processing, several automatic processes can run in parallel (ibid). In contrast, a controlled process utilizes a sequence of nodes controlled and temporarily activated in working memory through attention of an individual (ibid). Because controlled processing requires a lot of effort and cognitive resources, controlled processes cannot operate in parallel and suffer a bottleneck effect.

That is, when we consciously attend to one thing, we have to block out other things. If several demands are present at the same time, we need to prioritize them, putting others in line while attending to one of them. Thus, controlled processes are capacity-limited.

It has to be noted that controlled processing and automatic processing are not dichotomous but rather are two ends of a continuum. According to theories of skill acquisition (e.g. Anderson, 1983; Anderson, Greeno, & Kline, 1981; Shiffrin & Dumais, 1981), controlled processes can be automatized through practice (Newell & Rosenbloom, 1981). In this way, complex cognitive skills, such as solving physics or mathematical problems, using a text editor, or using a language, develop (McLaughlin, Rossman, & McLeod, 1983; Newell & Rosenbloom, 1981). In a novel context or in the face of a problem encountered during automatic processing, we engage in controlled processing, letting our central executive control the processing task. When controlled processes are repeated frequently and consistently enough, sequences of information processing will be transferred from working memory to long-term memory and a permanent learning of information may take place. Long-term storage of information is caused by associations made between previously separate elements through controlled processing (Shiffrin & Schneider, 1977).

Shiffrin and Schneider (1977) also discussed how the amount of effort in performing cognitive tasks can be reduced. Two significant factors were discovered: the degree of attention involved in doing the task and the amount of rehearsal with the material in the task. They found that the more attention a task requires, the more energy is consumed and the slower the processing. Conversely, the less attention a task demands, the less the amount of energy consumed and the faster the processing. The more familiar an individual is with the information or task material, the less the amount of attention required and the faster the processing. The more rehearsal with the task material one has had, the faster one will be able to process the information.

To sum up, a cognitive task, which requires a large amount of attention and effortful mental operations, involves greater controlled processing. A task, which requires little

attention and little energy, involves more automatic processing. Consistent practice of controlled processing will lead to automatic processing. In other words, in skill development, controlled processing is a stage one has to go through to achieve automatic processing. Automatic processing is a result of prolonged rehearsal of task material. It is a learned process and achieved only after a set of processing procedures has been built up through practice.

L2 learning is seen, from an information processing perspective, as a process of changing from controlled processing to automatic processing (Hulstijn, 1990; Karmiloff-Smith, 1986; McLaughlin, 1987). As stated above, automatic processing results from prolonged rehearsal of task material and this is where planning becomes important. If implemented sensitively and systematically over time, planning provides learners with multiple opportunities to rehearse material and thus may enable them to access such material through automatic processing.

Shiffrin and Schneider's ideas of controlled processing and automatic processing (1977) have provided explanations for a wide range of human behaviors. However, they do not specify how cognitive skills develop. This gap is filled by Anderson's model of human cognition (1983), Adaptive Control of Thought (referred to as ACT\*<sup>3</sup>), which the discussion will now turn to.

### **2.2.5 Anderson's ACT\***

Anderson's model aimed to provide an account for "the full range of skill acquisition: from language acquisition to problem-solving to schema abstraction" (Anderson, 1983, p.255). In this model two types of long-term memory are differentiated: declarative memory and procedural memory. Declarative memory stores information that can be consciously recalled, such as facts or encyclopedic knowledge of the world. In contrast, procedural memory is the memory for how to do things. It can be accessed without conscious control or attention. When needed, it is retrieved automatically and utilized for

---

<sup>3</sup>This is an extended version of the ACT theory (Anderson, 1978).

execution of behaviors.

According to Anderson, skill development goes through several stages: from the initial stage of application of declarative knowledge to proceduralization through synthesizing processes, during which sub-procedures of general-purpose production are combined and continual refinements are made (ibid). In the initial stage, declarative knowledge about how to perform a skill has to be retrieved from long-term memory and kept in working memory to be used interpretively for guiding behavior. Although flexible, it is a slow process and puts a great burden on working memory. It is slow because an individual has to interpret each unit of information in a series of small steps. This stage corresponds to the “controlled processing” stage in Shiffrin and Schneider (1977). In the second stage, declarative knowledge is converted to procedural knowledge through the knowledge compilation process by which knowledge keeps being restructured and new procedures are established. Compilation is further divided into *composition* and *proceduralization*. During *composition*, sub-procedures for solving a problem are combined into a single procedure, producing a speed-up effect (ibid). Then the new procedures undergo continual restructuring and fine-tuning, during which process generalizations (e.g. inducing rules) or discriminations (i.e. narrowing the scope of rules) are made, thus strengthening the procedures. This is the *proceduralization* process. When skill is proceduralized and becomes autonomous, information processing is sped up. “Automatic processing” in Shiffrin and Schneider’s model (1977) is then achieved.

In the course of proceduralization, practice plays an essential role. Through relevant practice over many trials, controlled processes gradually withdraw during performance and automatic processes take over. However, practice will cease to yield large returns in terms of improvement at some point, because optimal performance has been achieved (N. Ellis & Schmidt, 1998). This is the power law of practice (Newell & Rosenbloom, 1981).

Anderson’s model (1983) has provided explanation for development of a full range of skills, including language acquisition, and has been applied in a number of SLA studies (Bialystok & Smith, 1985; Hulstijn & Hulstijn, 1984; McLaughlin, 1987; Schmidt,

1992; Towell, Hawkins, & Bazergui, 1996). In general, the role of practice in L2 development has been recognized. How to practice to achieve effects has become a central issue in SLA research. Planning studies have attempted to address this issue by examining language produced under different conditions (i.e. with planning vs. without planning). However, up to now, the studies have only given snapshots of what it is like producing language under different conditions. The question of whether practice under planning conditions would facilitate proceduralization and automatization has not been explored. This is a gap the present study attempts to address.

## **2.3 Information processing and language learning**

In this section, the information-processing view of language acquisition will be discussed in relation to the three stages of information processing (i.e. input, central processing, and output). Before discussing language studies that centered on issues relating to the three stages of information-processing, general views on language skill acquisition and second language learning from an information-processing perspective will be introduced briefly.

### **2.3.1 Information-processing perspective of second language learning**

McLaughlin et al. (1983) were some of the first scholars to propose an information-processing approach to explain second language learning. In this approach, L2 “learning at the initial stage is seen to involve controlled processes with focal attention to task demands. As the learner becomes more familiar with the situation, attention demands are eased and automatic processes develop, allowing other controlled operations to be carried out in parallel with automatic processes as performance improves” (ibid, p.142). In the process of automatization, restructuring, which refers to replacing the old procedure with a more efficient one by integrating or reorganizing the components of a task into new units, occurs (McLaughlin, 1990). It is a process learners undergo as they modify their internalized cognitive representations.

In a similar vein, Skehan (2000) explains how interlanguage develops in terms of two stages: changing the underlying interlanguage system (i.e. restructuring) and gaining control of the system (i.e. automatization). At the former stage, interlanguage is extended and complexified through noticing, the role of which in L2 learning will be discussed in the next section. The underlying system of interlanguage is restructured and the new linguistic knowledge (gained through noticing) is integrated into coexistence with prior knowledge. At the stage of gaining control of the system, the new forms are becoming more target-like and a certain degree of automatization is achieved. Thus learners become more accurate and fluent in their performance (ibid). This view of IL development accords with McLaughlin's theory but specifies more what takes place at different stages of IL development. Therefore, it provides a basis for the measurement of change in interlanguage which has been operationalized in many studies, as will be discussed in Chapter 3.

Also within the framework of information-processing, but drawing more on theories about skill development (see Anderson's ACT\* in section 2.2.5), skill acquisition theories seek to explain language development from the perspective of change in knowledge representation and executive control of language skills. They see language development as a process of changing declarative knowledge, which involves controlled processing, to procedural knowledge and the subsequent gradual refining of it (Hulstijn, 1990; Karmiloff-Smith, 1986; McLaughlin, 1987). That is, development goes through three stages: declarative, procedural, and automatic (DeKeyser, 2007c). After the initial stage of acquiring declarative knowledge comes the proceduralization stage, in which declarative knowledge is turned into procedural knowledge, which has been 'programmed' into ready-made chunk available for use when called upon in real-time communication (similar to Anderson's *composition*). Once knowledge is proceduralized, it needs to be refined to make the use of it completely fluent and spontaneous (similar to Anderson's *proceduralization*). In other words, it needs to be automatized. In the gradual process of automatization, practice plays a central role. Through practice, processing

speed increases and error rate decreases.

Then the central issue is what kind of practice will lead to proceduralization and automaticity. DeKeyser (1998) contends that practice should be directed at ‘behavior’ (i.e. develop automatic processes). As he explains, “[...] proceduralization is achieved by engaging in the target behavior—or procedure—while temporarily leaning on declarative crutches [...] Repeated behaviors of this kind allow the restructuring of declarative knowledge in ways that make it easier to proceduralize and allow the combination of co-occurring elements into larger chunks that reduce the working memory load” (ibid, p.49). This means that pedagogical measures taken to help learners produce language in a more target-like manner would facilitate proceduralization. It has been generally agreed that good practice should provide opportunities for comprehending or expressing real thoughts (DeKeyser, 2007a). How this ‘good practice’ can be operationalized and implemented in a language classroom is the central concern of task-based research. As will be seen in the next chapter, research on tasks and task conditions has shown that different degrees of planning, time pressure, repetition, etc. have impact on accuracy, fluency, and complexity (three aspects of skill development) of learners’ performance. These task conditions may favor or disfavor one or more of these three aspects. That is task condition may play a role in facilitating L2 skill development in these three aspects. However, so far, this area of research has focused on immediate task performance and the findings are somewhat inconsistent. This makes the insight from this line of research rather limited.

Summarizing the information-processing views of language acquisition discussed above, language development entails qualitative changes in knowledge representation (the concept of restructuring in Skehan’s account and proceduralization in skill acquisition theories) and in control over that knowledge, through automatization process in which practice plays an essential role. With this understanding of language development, it is not hard to see a role for planning to play in the process. First, planning has the potential of inducing more complex language, which is needed for expression of

complex ideas generated through planning. In this way, the restructuring process is triggered. It might also facilitate automatization in that planning enables access to the not yet fully automatized knowledge and through repeated practice it will become automatic.

Having presented the information-processing view of language learning, the discussion will now turn to the SLA studies relating to the three stages of information processing.

### **2.3.2 Input**

As stated earlier, not all input in the environment can be picked up and processed. Attention is needed for information to be picked up and enter the central processing unit. In light of this view, several researchers in second language acquisition (SLA) have stressed the role of attention in the input stage (Gass, 1997; Schmidt, 1990, 2001; VanPatten, 1990). In SLA, input refers to the language learners are exposed to, whether in written form or aural form. Schmidt (1990) points out the importance of noticing in second language learning. In line with the information processing theory, Schmidt argues that not all input has the same value and only the noticed input can become intake and be effectively processed. Moreover, he contends that the more frequently a form occurs in the input and the more salient it is, the more likely it is to be noticed and later incorporated in the L2 system.

Recognizing the important role of noticing, VanPatten (1990) demonstrates what learners tend to be attracted to in the input. He claims that learners process meaning in the input before they process the form. For this reason, they pay attention to content words more than anything else; they prefer to process lexical items more than grammatical items; and they process meaningful morphology more than less meaningful morphology. They would attend to non-meaningful forms only when they could process the communication content with no or very little effort. Based on these findings, he proposed an input processing approach to language teaching (VanPatten, 1996), in which learners are trained to process input in a way that makes them more able to notice the



cues in the input so that the links between meaning and form could be established.

### **2.3.3 Central processing**

At the central processing stage, the operations of the memory system (working memory and long-term memory) play a crucial role in processing the input information. Working memory is the workspace where input is extracted from the environment and processed with the contextual knowledge from long-term memory activated for processing. Later, the processed information may feed into long-term memory and the knowledge base in long-term memory is restructured. Applying this to language learning, language knowledge stored in long-term memory is restructured through the action of working memory. That is, the interaction of working memory and long-term memory plays a crucial role in the development of language.

In cognitive psychology, there have been arguments over representations in long-term memory. Some scholars contend that long-term memory consists of rule-based systems whereas some others argue for exemplar-based systems. Proponents of the rule-based systems (e.g. Reber, 1989) assume that learning involves induction from the stimulus material and generalization of abstract rules. Development of this system is indicated by growth or increase in complexity of the underlying system. Advocates of the memory-based system (or exemplar-based system) argue that what is learned is the accumulation of ready-made chunks (Carr & Curren, 1994; Schmidt, 1995). Instead of relying on analyzing rules, learners use the exemplar-based system to match the current input with the previous input known to be correct. Though the exemplars may contain structural rules, they are learned as chunks (Carr & Curren, 1994). Still some other scholars (e.g. Carr & Curren, 1994; Mathews et al., 1989) believe that there is a dual mode of processing in long-term memory and both rule-based learning and exemplar-based learning are evidenced. Based on the findings outlined above, Skehan (1998) proposes that in an L2 system “long-term memory consists of a rule-based analytical system, a memory-based formulaic system, and general schematic knowledge”

(p.58). The rule-based system and the exemplar-based system coexist and language learners naturally move from one system to the other. The three components of long-term memory are assumed to interact through the mediating actions of working memory (Skehan, 1998). Working memory is the basis for allocation of attentional resources and is the area where ongoing language processing takes place. Through the actions of working memory, the knowledge base in long-term memory is changed with the rule-based system being extended or restructured and the memory-based system increased in stock of exemplars.

Working memory and long-term memory are the mechanisms at work in the process of L2 development outlined above. How they work particularly at the output stage will be discussed next.

### **2.3.4 Output**

The way working memory and long-term memory work at the output stage could be briefly, maybe simplistically, described in the following way. Compared with L1 production, L2 production is more effortful and less automatic. In this case, materials relating to both content and the linguistic forms of the message from long-term memory need to be retrieved. The retrieved materials are stored temporarily and organized in working memory. Through the interaction of working memory and long-term memory, language is produced. With practice in producing language, controlled processing gradually becomes automatic and the representation of knowledge in long-term memory is restructured. Seen from this perspective, output can be considered as being facilitative to language learning. This view is explained explicitly by the Output Hypothesis, which will be discussed in the next section.

#### **2.3.4.1 The Output Hypothesis**

The Output Hypothesis proposes that through producing language, either spoken or written, language learning may take place (Swain, 1993). This hypothesis was advanced

by Swain (1985, 1993, 1995, 1998) based on her decades of study on Canadian Immersion Programs. Research had shown that providing learners with comprehensible input only could not lead to a high level of L2 accuracy. After years of study, students in those programs still lacked grammatical accuracy though they gained high level listening comprehension skills and communicative fluency. Due to this observation, Swain (1985) argued that output is not only a demonstration of what is acquired (Krashen, 1989) but a necessary step towards acquisition. For her, output serves several functions in relation to language acquisition. First, it enhances fluency because it offers learners opportunities to practice the language. It is argued that through meaningful practice learners' production processes may be automatized, resulting in increased fluency (Swain, 1993). In this respect, the Output Hypothesis is in line with the information-processing perspective of L2 learning discussed previously, which contends that practice will result in a decrease not only in reaction time (i.e. an increase in fluency) but also in error rate (i.e. an increase in accuracy) (DeKeyser, 2007b, 2007c; Skehan, 2000). Swain also claims that (1993, 1995), output has the function of enhancing accuracy in the following ways: 1. It promotes noticing. That is, in production learners may "notice the gap between what they want to say and what they can say" (1995, p.126). When learners notice the gap, they may search their own linguistic repertoire for information, seeking to close the gap by generating new knowledge or consolidating their existing knowledge. Output also helps them recognize their linguistic problems, thus motivating them to seek input that would solve their problems. 2. It lends opportunities for learners to test out their hypothesis about the comprehensibility and well-formedness of their language. A learner's output is an indication of hypotheses s/he has formulated about how a language works and that s/he is testing out. If the hypothesis is disconfirmed by negative feedback, the learner would reformulate these hypotheses leading to restructuring of his/her interlanguage, which in turn leads to interlanguage development. 3. It serves a metalinguistic function, giving learners a chance to reflect on their own language use. Under certain task conditions, learners use language to talk about language form. The content of negotiation

is linguistic forms and their relation to the meanings they are trying to express. In so doing, they are enabled to control and internalize linguistic knowledge.

As pointed out at the beginning of this chapter, planning is an activity naturally involved in language production (i.e. output). It affects the way learners allocate their attention during production and has the potential of enhancing language performance, as will be more evident in the following sections and Chapter 3. Thus, it follows that producing language output, spoken or written, under planning conditions may promote IL development.

After outlining the importance of output in language learning, the discussion will now turn to the processes of producing output.

#### **2.3.4.2 Speech production model**

One influential model of speech production that has guided production-related studies in SLA is Levelt's speech production model (1989), which describes the process of speech production in L1. Levelt applies information processing to language production, positing that language production is constrained by the same factors that regulate other cognitive activities. He divides the speech production system into three principal processing components: a conceptualizer, a formulator, and an articulator. The conceptualizer decides the intention of the speaker and selects information that is relevant. During the conceptualizing stage two levels of planning take place. At the macro-planning level, the communicative goal is elaborated into subgoals and relevant information for realizing the subgoals is retrieved. At the micro level, details of the structure of the message are established, resulting in a preverbal message. This preverbal message is then sent to the formulator. The formulator encodes the preverbal message into grammatical and phonological features. The lexicon is also accessed for words to be used. The formulator produces a phonetic plan that is then forwarded to the articulator. The articulator converts the phonetic plan into overt speech. During the process of production, there is a self-monitoring mechanism regulating the three production stages.

It inspects the preverbal message to see whether it matches the speaker's original intention. It also inspects the phonetic plan before it is articulated and monitors overt speech. Corresponding to an information processing framework, the speech production process can be described as follows. During the stage of conceptualizing, propositional contents in the long-term memory need to be accessed and the linguistic forms necessary for conveying the message are extracted at the formulating stage. This information is orchestrated in working memory and then sent to the articulator.

Levelt's model (1989) also gives an account of sources of variation from a psycholinguistic perspective. In the 'conceptualizer', decisions are made regarding the variety of language to use according to factors such as the social contexts and the established communicative goals. At the formulation stage, where ideas are translated into a speech plan, a conscious decision has to be made on the style of language to be used. Appropriate words, sentence structures and phonological rules are selected for use. In the 'articulator', the speech plan is executed into actual speech. This model suggests that speech variability is determined by social and linguistic contexts. In speech production, accessing words in the lexicon, grammatical encoding, and assigning phonological encoding are dimensions competing for mental effort.

Levelt's model has been adapted to account for bilingual speech production (De Bot, 1992; Kormos, 2006) without undergoing substantial changes. In L1 production both controlled processing and automatic processing are involved. The conceptualizing and monitoring processes operate under controlled processing whereas the formulating and articulating run automatically. In the case of L2 speech production, however, because of their limitation in L2 knowledge and a heavy dependence on declarative knowledge when producing speech, learners are more likely to have to execute formulation and articulation with controlled processing and thus experience difficulty in those two stages of production (De Bot, 1992). Thus, compared with L1 production, L2 production involves only partial automatic processing and requires more attention (Kormos, 2006).

While the model of speech production provides insights into processes of producing

output, oral output in particular, models of writing processes are more relevant to the present study, since it investigates the effects of manipulation of one component of writing processes (i.e. planning) on written output. Therefore, the following section will be devoted to a discussion of models of writing.

#### **2.3.4.3 Models of writing**

Theories on writing (e.g. Bereiter & Scardamalia, 1987; Grabe, 2001; Hayes & Flower, 1980; Hayes & Nash, 1996; Kellogg, 1996; Zimmerman, 2000) posit that writing undergoes processes similar to those of speech production in general (i.e. from conceptualizing, formulation, to producing output). However, there are fundamental differences between the two modes of production. For example, writing usually takes a longer time than speaking and leaves a physical trace that can be referred to later by the writer or the reader (Grabowski, 1996). The two modes are also different in terms of the demands put on cognitive resources. During the writing process, the addressee is not present and this has important cognitive implications. Because of the absence of the addressee, writers do not have to use cognitive strategies to maintain the flow of conversation. Therefore, they are under less communicative pressure, which allows them more time to engage in planning and information retrieval. On the other hand, the absence of the addressee also presents difficulty for the writers since they have to anticipate the audience and manage information in a way that helps the reader understand the intended meaning. Given these differences, it will come as no surprise that differences appear in descriptions of writing process.

Several researchers have proposed models of writing and one often cited model in the literature on writing process is that of Hayes and Flower (1980). There are three main components in this model: task environment, writer's long-term memory, and a number of cognitive processes, including planning, translating ideas into text, and revising. Task environment refers to the writing assignment and the text produced so far and writer's long-term memory includes knowledge of topic, knowledge of audience, and stored

writing plans. The most important insight this model brought to the understanding of writing process is that writing is a recursive process, which involves continuous backward and forward movement between the already written text and the emerging text. Since it was proposed, this model has generated a lot of discussion and review. One of the criticisms is that it does not take into consideration influential factors, such as the social context, the motivation, and task variation (e.g. Grabe, 2001).

In 1996, Hayes updated the model and the new model consists of two major parts: the *task environment* and the *individual*. In the new model, *task environment* is redefined as including the social environment and the physical environment. In fact, the task environment was not very much different from that of the old model in that the social environment refers to the writing assignment and the physical environment refers to the text produced so far, even though the elements in each of the sub-component differ a little from those in the old model<sup>4</sup>. The *individual* aspect in this model includes four interacting components: *working memory*, *long-term memory*, *motivation and affect*, and *cognitive processes*. Hayes's model of working memory is based on Baddeley's account of working memory (1986) with some adaptations. In Hayes's model, *working memory* consists of a phonological memory (phonological loop in Baddeley's term), which stores auditory/verbal information, a visual-spatial sketchpad, which holds visually or spatially coded information (e.g. written words or graphs), and a semantic memory, which stores conceptual information. *Long-term memory* includes task schemas, topic knowledge, audience knowledge, linguistic knowledge, and genre knowledge. Task schemas are defined by Hayes as "packages of information stored in long-term memory that specify how to carry out a particular task" (Hayes, 1996, p. 24). Topic knowledge refers to what the writer knows about the topic while audience knowledge refers to the writer's consideration for the readers. Linguistic knowledge includes knowledge about the language resources that are needed in the writing process whereas genre knowledge is the knowledge about how to write for a particular genre required by a task.

---

<sup>4</sup>Readers are referred to Hayes (1996) for a chart displaying all the sub-components in the model.

The cognitive processes in his model include *text interpretation*, *reflection*, and *text production*. *Text interpretation* refers to the process of creating mental concepts from listening to, reading, or scanning linguistic and graphic input. *Reflection* creates new information by combining existing information with interpretation of the new input. Finally, new linguistic (written or spoken) or graphic output is produced from the mental concepts in the text production stage. The cognitive processes interact with both working memory and long-term memory. Obviously, Hayes's model draws upon information processing theory, with text interpretation, reflection, and text production corresponding to the input, central processing, and output stages of information processing respectively.

Hayes recognizes the influence of motivation and affect on writing. The motivation and affect component in the model includes goals, predispositions, beliefs and attitudes, and cost/benefit estimates. All of these factors may influence the way a writer goes about the writing task and the effort s/he will put into the task.

Hayes's model is significant because it provides a comprehensive description of the various factors that influence writing, particularly in terms of the roles of working memory, long-term memory, cognitive processes, and motivation/affect. However, the model has shortcomings in several aspects. For example, it lacks specificity in accounting for important situational factors (Weigle, 2002), such as the settings (e.g. classrooms, libraries, or computer labs), the task (e.g. letter writing, essay writing), and topic for academic writing. While recognizing the important role of working memory in the process of writing, it does not give a detailed account of how working memory works and how it interacts with other components in the writing model.

Kellogg proposed a model of working memory in writing (Kellogg, 1996) that fills the gap mentioned above. This model sees formulation, execution, and monitoring as three systems involved in text production, under each of which are two sub-processes. Figure 1 shows Kellogg's model.



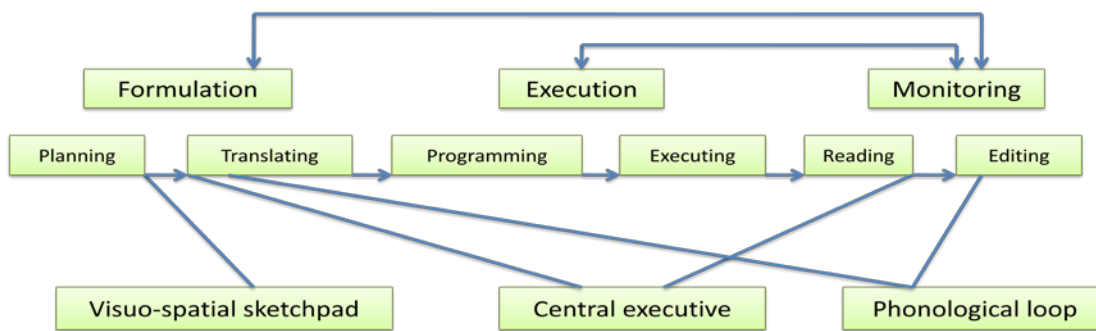


Figure 1 Kellogg's model of working memory in writing

Planning ideas and translating them into intelligible sentences are two processes in formulation. The results of planning are the sources for translating and then the output of translating is sent to programming in the execution system. The outputs of formulation are, at the same time, forwarded to editing in the monitoring system before handwriting or typing takes place. This model does not imply that writers follow a linear processing procedure from formulation of ideas to execution and then to monitoring. Instead, it allows simultaneous activation of the three systems as long as the demands on central execution do not exceed the capacity limitations. The three systems interact with one another. For example, editing takes place before the execution of motor movements and it may also cause reformulation. The execution of a word or sentence may take place simultaneously with planning and translating of new ideas or monitoring of written material. Obviously, in Kellogg's model planning is considered an important step in writing.

It has to be noted that in research on writing, unlike research on speech production, planning is a term referring to goal setting and ideas generating only whereas in speaking planning may include converting ideas into words and sentences. The job of converting ideas into words and sentences is termed translation in the literature on written production. It is commonplace for ideas to be translated partially as planning is going on. This corresponds to what Vygotsky described as 'inner speech' (Vygotsky, 1962), which is "to a large extent thinking in pure word meanings. It is a dynamic, shifting, unstable thing, fluttering between word and thought." (p. 149). Therefore, in the present study pre-task planning involves idea planning and partial translation of ideas. Some of the

translated ideas were written on planning sheets while some others were kept in participants' minds.

While processes of writing described in Hayes's model of writing can find corresponding stages in information processing discussed earlier, Kellogg's model concentrates on the output stage with a focus on the function of working memory in writing production. Kellogg's model presumes that the formulation system places the heaviest burden on working memory. To begin with, generating ideas, thinking of how to organize them into appropriate discourse, and choosing the right tone are tasks placing demands on working memory capacity. Therefore, planning could be thought of as engaging the central executive and drawing heavily on the limited memory capacity. Translating ideas into intelligible sentences also demands resources of the central executive when the writer has to struggle to find the right words and sentence structures to express certain meaning. Compared with execution demands, formulation demands on central capacity take priority and have to be satisfied first. This is supported by a study (Brown, McDonald, Brown, & Carr, 1988), which showed that when formulation demands on memory retrieval increased, legibility decreased and errors increased. As far as monitoring is concerned, editing places heavy demands on the central executive. Monitoring suffers when formulation demands increase.

The key feature of Kellogg's model is that working memory has limited capacity and for this reason the writer has to make decisions on which process to prioritize when under pressure to produce a piece of writing. Allocation of attention impacts on the quality of writing. In Kellogg's model of writing process, it is easy to see how task planning can have an effect on writing performance. If a good deal of idea planning and partial translation is taken care of before a task, the writer would be able to focus more on execution of plans and monitoring. As a result, the online processing burden on working memory is lessened and an improved writing performance can be expected.

Because of the nature of planning and its expected impact on language production outlined above, it has important implications for language teaching. A recent pedagogical

development in second language teaching is task-based language teaching and learning (e.g. R. Ellis, 2003; Samuda & Bygate, 2008; Skehan, 1996a; Willis & Willis, 2007). In this approach to language teaching, planning is a task implementation condition manipulable for a teacher to increase or decrease task demands, depending on the objectives of the course or class. The following section will briefly review the task-based approach to language teaching and learning (TBLT).

## **2.4 Task-based language teaching and learning**

TBLT is the pedagogical context, in which most planning studies are carried out. A discussion of the main concepts and issues in TBLT will enable a better understanding of why planning has been studied and the value of this line of research. In the following sections, the definition of ‘task’, the issue of task complexity, and task implementation will be discussed, after a brief introduction to TBLT.

In light of an information processing view of language acquisition (e.g. skill acquisition theories), declarative language knowledge can be converted to procedural knowledge through practice. This suggests a facilitative role of instruction in second language learning. That is, learners can obtain declarative knowledge through instruction then proceduralize it through practice, during which process their L2 knowledge and skills are gradually restructured and fine-tuned, leading to L2 development. Then comes the question of what kind of instruction is effective. A current approach to language teaching is TBLT (a version of communicative language teaching (CLT)). Different from the more traditional approaches, in which language is seen as the object of teaching therefore taught and exercised before learners are required to use it to communicate, tasks are used in TBLT. When doing tasks, learners are engaged in communicating meaning, during which process learners are “using English to learn it” (Howatt, 1984, p. 279). This is out of the belief that language learning does not involve only learning linguistic knowledge (e.g. grammar, phonology, and lexis) but also learning how to use it in real communication---a belief underlying CLT. Therefore, learning through communication is

thought to be more effective.

Such an approach of language teaching is influenced and supported by several key theories in SLA. For example, Krashen's Input Hypothesis (Krashen, 1981, 1982) has it that exposing learners to meaningful input that is just slightly beyond their current linguistic competence but still comprehensible for learners would lead to development of their L2 system. That is, comprehensible input provides an environment, in which language is acquired subconsciously. In this way, learners acquire incidentally the formal aspects of language while their focus of attention is on meaning. Another separate but related theory is the Interaction Hypothesis (Long, 1983, 1996). According to this theory, conversational modifications (e.g. request for clarification, confirmation checks) learners make during meaning negotiation make input comprehensible. Moreover, the negative feedback obtained during negotiation for meaning may orient learners to notice the 'gaps' and 'holes' (cf. the Output Hypothesis) in their current interlanguage. TBLT can obviously find support from the theories discussed above. Tasks provide learners with opportunities to engage in meaning comprehension or production, during which process comprehensible input is available and learners may learn the language incidentally. Production tasks can create opportunities for meaning negotiation which in turn promotes L2 development.

Research in SLA has also found that focusing on meaning only also presents problems. As Skehan (1996a) pointed out, when negotiating for meaning, learners can rely on formulaic expressions and/or communicative strategies, such as using body language and intonation, instead of using appropriate linguistic forms to express meaning. This might lead to fossilization (i.e. a stage at which L2 progress ceases). Therefore, it is generally agreed that effective instruction needs to create opportunities to 'focus on form' (Long, 2000; Long & Crookes, 1991; Long & Robinson, 1998), which in the context of language pedagogy refers to the instructional intervention to direct learners' attention to the formal aspects of language while their primary attention is on meaning. As discussed in Section 2.2.3, human working memory capacity is limited. Because of this, it is

difficult for learners to attend to both meaning and form simultaneously when trying to comprehend input (e.g. VanPatten, 1990) or producing output. Learners tend to prioritize meaning in communication, overlooking some linguistic features, especially those that are non-salient and redundant. Thus, measures need to be taken to draw learners' attention to form. There is recognition of this need in TBLT. Learner's attention to form is usually drawn in the post-task phase, but, in the case of planning, it can happen in the pre-task phase as well. During planning, learners can pay attention to linguistic forms that have arisen out of their need to communicate meaning. This will be elaborated in 2.4.3. For now, the discussion will turn to the definition of 'task'.

#### **2.4.1 Definition of 'task'**

In the past two decades, tasks have been defined in a variety of ways (e.g. Candlin, 1987; R. Ellis, 2003; Long, 1985; Prabhu, 1987; Skehan, 1998). Though these scholars summarize characteristics of language tasks in different ways, there is a lot of common ground shared in these definitions. For example, Skehan (1998, p. 95) defines 'task' as an activity in which

- Meaning is primary.
- There is some communication problem to solve;
- There is some sort of relationship to comparable real-world activities;
- Task completion has some priority;
- The assessment of the task is in terms of outcome.

Ellis's (2003, p. 9-10) definition of language tasks is as follows:

- A task is a workplan;
- A task involves a primary focus on meaning;
- A task involves real-world processes of language use,
- A task can involve any of the four language skills;
- A task engages cognitive processes;
- A task has a clearly defined communicative outcome.

In both definitions, a task is seen to serve communication needs with a primary focus on meaning. Another feature of a task depicted in both definitions is that there is some connection between task activities and real world language use and that a task is assessed in terms of communicative outcome.

In many studies on task-based language teaching, tasks are operationalized as communicative oral tasks only. Thus, research in this field has focused more on oral tasks. However, as Ellis (2003) pointed out, ‘a task can involve any of the four language skills’ (p.10). There is a need for more studies to be conducted in a writing context to achieve a more comprehensive understanding of TBLT.

#### **2.4.2 Task complexity**

In this thesis, the term task complexity refers to essential components and characteristics of tasks, such as information cues, required acts, and expected products, which differentiate simple tasks from complex tasks. The issue of task complexity arose out of a need for establishing criteria for sequencing tasks (e.g. from simple to complex and from easy to difficult) in a task-based syllabus. It has been discussed and measured with different frameworks. The Cognition Hypothesis (Robinson, 1995, 2001, 2003, 2005) and the Limited Capacity Hypothesis (Skehan, 1996a, 1998) are two most well-known models.

In Skehan’s view (1996a, 1998), there are three components to task complexity: code complexity, cognitive complexity, and communicative stress. Code complexity is concerned with the linguistic demands of a task; cognitive complexity has to do with the content of the task and how it is structured; and communicative pressure refers to the performance conditions, such as time pressure, modality, and control.

In terms of the impact of task complexity on task performance, Skehan (1996a, 1998), arguing from an information processing view, posits that trade-offs between fluency (i.e. the ability to produce language at normal rate without undue pauses), accuracy (i.e. correctness of language), and complexity (i.e. elaborateness) should occur in language

production because learners' processing capacity is limited. Attention to one aspect of language will cause other aspects to suffer. As has been mentioned in Section 2.3.4, Skehan (1998) distinguishes an exemplar-based system and a rule-based system in L2 system. The exemplar-based system includes lexical items and ready-made formulaic chunks whereas the rule-based system, which requires more processing in production, is composed of language rules. Skehan (1998) suggests that learners' language varies according to which aspect of language they choose to focus on. When fluency is prioritized, as in real time communication where communicative needs are pressing, learners draw on their exemplar-based system and rely on the ready-made formulaic chunks of language. When learners need to engage in planned discourse or have an opportunity to focus on linguistic forms, as in the case where they could plan prior to task performance, they are more likely to access their rule-based system and produce more accurate and complex language.

Also from an information processing perspective, Robinson (2001, 2005), sees task complexity as "the result of the attention, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner. These differences in information processing demands, resulting from design characteristics, are relatively fixed and invariant" (Robinson, 2001, p.28). According to his Triadic Componential Framework for task classification and task design, there are three broad categories of task features contributing to cognitive demands a task places on learners: task complexity, task conditions, and task difficulty. Task conditions concern interactive factors, such as if the task requires one-way or two-way interaction. Task difficulty concerns how difficult the task is as perceived by the learners. Regarding task complexity, he distinguishes two task complexity dimensions: resource-directing and resource-dispersing. Resource-directing elements refer to task characteristics that determine the concepts required by the task to convey. A task that requires learners to engage in causal reasoning, for example, is more complex than one that does not have this requirement. These task variables are considered resource-directing because the cognitive or conceptual demands

they put on learners direct their attentional and memory resources to their L2 system so that they can accurately understand and convey the concepts (Robinson & Gilabert, 2007). The more resource-directing variables there are the more complex a task is. Resource-dispersing elements include task variables that provide favorable task conditions, such as providing planning time or background knowledge for the task. These variables do not direct learners' attention to their L2 system but rather facilitate automatic access to and control of their existing L2 knowledge. Tasks without these supporting elements are considered more complex than those with these supporting elements.

Robinson (1995, 2001, 2003, 2005) argues that L2 learners are capable of parallel processing and can attend to more than one aspect of language simultaneously. He claims that increasing cognitive demands, which result in making a task relatively more complex, would push learners to produce language with greater accuracy and complexity. He believes that language complexity is determined by the cognitive complexity of the task and there is no competition between accuracy and complexity. In general, Robinson (2001, 2005) argues that increasing task complexity along the resource-directing dimension should result in more accurate and complex production whereas increasing task complexity along the resource-dispersing dimension, by removing task supportive elements, such as planning time, would lead to a decrease in accuracy, complexity, and fluency.

From the brief review of the two models outlined above, it can be seen that Skehan and Robinson make the same predictions for the effects of planning time. In Robinson's own words, "where the Cognition Hypothesis differs from the Limited Capacity Hypothesis is over the claims [described above] for the beneficial effects on accuracy and complexity of increasing the resource-directing dimensions of tasks. The resource-directing/dispersing distinction is one that Skehan does not make, leading him to claim that complex task performance, along any dimension, degrades accuracy, fluency, and complexity simultaneously" (Robinson & Gilabert, 2007, p. 167). However, another difference between the two theories is that a trade-off in the three aspects of language performance is predicted in the Limited Capacity Hypothesis (Skehan, 1996a, 1998)



whereas such a prediction is not present in the Cognition Hypothesis (Robinson, 1995, 2001, 2005).

The present study does not set out to test the two theories but will draw on them in understanding the results of the study, in particular to explain how task variables influence the results.

### 2.4.3 Task implementation

As has been pointed out earlier, SLA research has found that comprehensible input, the opportunities to engage in interaction, and to focus on form all have important roles to play in promoting L2 development. TBLT may create all these opportunities if implemented carefully and systematically.

Drawing on SLA research and information processing theories, Skehan (1996a) proposed a framework for implementation of TBLT that allows techniques to be used to balance the attentional focus on meaning and form. This framework is shown in the figure below.

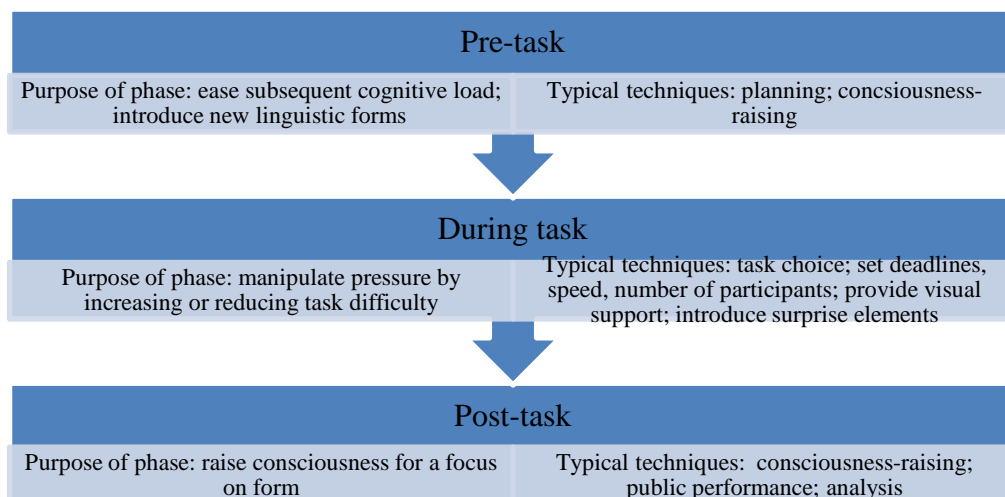


Figure 2 Stages in task implementation (modified based on Skehan, 1996a, 1996b)

In this framework, Skehan shows that the cognitive overload during task caused by limited attentional capacity could be eased by techniques employed at the pre-task stage. Planning is one of these techniques that allow learners to prepare cognitively and linguistically for the subsequent task, thus having the potential for releasing more

attention for the actual language that is used during task completion. The result might be increased accuracy, complexity, and/or fluency in task performance (see Chapter 3 for the effects of planning). At the pre-task stage, a language focus also takes place, when task-related linguistic forms are introduced. Though the new forms might not be learned in the short term, this focus on language may come to fruition sometime in the future.

At the during task stage, various techniques can be used to manipulate task difficulty. For example, communicative pressure for task completion could be increased by placing a time limit. Task modalities (i.e. speaking, writing) also differ in communicative pressure they impose on learners (Skehan, 1998). Difference in pressure may influence the amount of attention available at this stage, and thus affecting learners' performance in accuracy, complexity, and fluency (Skehan, 1998). At the last stage, learners' attention to form can be drawn again to promote pedagogic goals. In this phase, focus on form can take place in several ways. For example, students could be given the opportunity to repeat a task while paying some attention to form during this second execution of a task that they already have some familiarity with.

If tasks are implemented in the way Skehan proposes, the danger of reaching exclusively the goal of completion that may result in prioritizing fluency at the expense of accuracy and complexity can be reduced. A balance between focusing on meaning and form may be achieved. Under the influence of Skehan's task implementation framework, the present study investigates the effects of pre-task planning (at the pre-task stage) and online planning (at the during task stage) on L2 writing performance and writing development.

## **2.5 Summary**

The information processing framework is the theory underpinning the present study. In other words, the present study accepts the assumptions made in information processing theory: 1. Human beings have limited capacity in terms of focus of attention and the ability to process information (Newell & Simons, 1972). 2. Tasks that are unfamiliar and

not well practiced require more attention and processing time than tasks that are familiar and well practiced. 3. Learning complex skills, such as second language learning, is a process of moving from controlled processing to more automatic processing (Shiffrin & Schneider, 1977). 4. Since different tasks require different amounts of time depending on the degree of familiarity with the task and prior practice of it, time could be taken as an index of the amount of effort needed for the task. In other words, provision of time could influence task performance. In the Information Processing Framework, the Skill Acquisition Theory (i.e. Anderson's ACT\*, 1983) is of particular interest for the present study, since it provides a detailed account of how cognitive skills develop. This theory claims that skills develop through conversion of declarative knowledge into procedural knowledge. Language skill is viewed as a complex cognitive skill and its acquisition follows the path from controlled processing to automatic processing through practice.

Applying information processing theories to L2 learning, L2 is considered to develop through a process of changing declarative knowledge to procedural knowledge and from controlled processing to automatic processing (Hulstijn, 1990; McLaughlin, 1987; McLaughlin & Heredia, 1996; McLaughlin et al., 1983). During this process, restructuring (McLaughlin, 1990) occurs and learners gradually gain more control of the reorganized language system (Skehan, 2000) through practice. In the account of L2 learning from an information processing perspective, scholars often give emphasis to the role of attention at both the input stage and output stage. At the input stage, only the input that has been attended to and noticed by the learner can become intake and later be effectively processed (Schmidt, 1990). Thus, noticing has been considered a necessary condition for L2 learning (ibid). At the output stage, propositional content and linguistic forms needed for conveying the message are first retrieved from long-term memory and stored and orchestrated in working memory for producing output. In working memory, attentional resources are allocated to attend to meaning or form. Through the actions of working memory, knowledge in long-term memory is restructured, resulting in a decrease in reaction time and error rate (DeKeyser, 2007c).

In line with this understanding, the Output Hypothesis (Swain, 1993, 1995) is advanced, in which it is explicitly stated that output is facilitative to language learning. The Output Hypothesis claims that output is not only a demonstration of what one has learned but provides an opportunity for processing language in a way that promotes language development. It is in itself part of the learning process. The Output Hypothesis could be considered as the cornerstone of the current study. This is because a premise for studying the relationship between planning and second language learning is to accept the assumption that production facilitates second language learning. This assumption has been verified by quite a number of empirical studies (e.g. Izumi, 2000, 2003; Izumi, Bigelow, Fujiwara, & Fearnow, 1999; LaPierre, 1994; Swain & Lapkin, 1995, 1998). Another assumption of the Output Hypothesis is that if learners are ‘pushed’ to produce language that is accurate and complex, their IL development can be fostered.

Task planning, as will be seen in the next chapter, may have the function of alleviating the cognitive load of a task and inducing focus on form, thus having the potential of helping learners produce more complex, more accurate, and/or more fluent language (Skehan, 1998). It could be assumed that if learners could repeatedly practice producing output under a planning condition, their process of proceduralizing L2 knowledge might be facilitated (DeKeyser, 1998), leading to IL development (Skehan, 2000). However, there has been very little empirical evidence to prove this assumption (R. Ellis, 2005b). The present study attempts to find out the relationship between enhanced output brought about by planning and L2 learning.

To understand better why and how planning can enhance output, models of language production have been reviewed, with a particular focus on models of writing because the present study concerns the impact of task planning on L2 writing. These language production models (Levelt, 1989; Hayes, 1996; Kellogg, 1996) provide a basis for understanding the processes by which learners engage in language production and the effects of planning on production. In models of writing, for example, planning is thought to be a sub-process of writing, in which ideas for writing are generated and language

necessary for expressing the ideas (at least partially) is formulated. Therefore, planning is an important process that could influence writing performance.

While theories outlined above provide a theoretical background for the present study, task-based language teaching provides the pedagogical background. It has close relevance to the present study for the obvious reason that the focus of the present study is the effect of task planning on L2 writing performance and its relationship with L2 development. So far, in the field of SLA a small body of research has been conducted on task planning with some significant findings, which provide a solid ground for the present study. These studies have been fruitful but on the other hand have left some questions unanswered. The next chapter presents a summary of these studies, making clearer the research gaps that still exist in this line of research.

## **CHAPTER 3 THE EFFECTS OF PLANNING ON LANGUAGE PRODUCTION: A LITERATURE REVIEW**

### **3.1 Overview**

With the popularization of task-based language learning and a growing interest in the role of attention and focus on form in SLA, the number of studies on planning has been increasing. Various variables in planning, such as length of planning time, task type, focus of planning, types of planning, and learners' proficiency level, have been investigated. However, as will be seen in later sections, most of the studies were conducted in oral contexts and only looked into the effects of planning on the immediate task performance. Because there has been no study, to date, that investigates the longer-term effects of planning on L2 writing development, the bulk of this chapter will be a summary of studies on the immediate effects of planning on L2 production.

In the following sections, the nature of planning, classification of planning, and how task-based language performance has been measured will be discussed as a precursor to a review of studies. Then the empirical studies on the effects of task planning on oral production will be reviewed (section 3.5). In sections 3.6 to 3.8, the impact of planning on written production is discussed. In these sections, the impact of planning on L1 writing and that on L2 writing are discussed separately. Section 3.9 is a summary of the review and 3.10 states the research gaps the present study addresses.

### **3.2 The nature of planning**

From a cognitive psychology point of view, the perspective of an information processing framework in particular, planning is viewed as a problem solving technique (Newell & Simon, 1972). It is a mental activity that involves goal setting, problem anticipation, information gathering and organization, and strategy development for the purpose of accomplishing a certain task (Simon, 1978). It is a process people naturally engage in when they set about solving a problem. Its nature and connections with

language production were pointed out by Das et al (1996):

“[Planning] is oriented toward the future and may include the creation and selection of problems, as well as the anticipation of a sequence of actions to solve them. Instead of being a hierarchical and linear process, planning is often nonlinear and revisionary in nature, and the formation and execution of a plan can occur simultaneously. Planning is a self-organizing, reflective process that the individual is conscious of at least at some point in its development and it requires motivation and metacognitive skills. It is an activity that integrates several different components and levels of functioning into one schema and is a uniquely human function with close connections to speech and language” (p. 54).

Planning, is thus regarded as a recursive process with multiple components and functions.

In studies on language learning, planning was traditionally seen as a synonym of monitoring (Krashen, 1981). It is associated in particular with language production but the activities involved in it are not exactly the same in different contexts. In studies on the process of speech production (e.g. Levelt, 1989), planning is differentiated from monitoring and only refers to the process of retrieving and organizing information before an utterance is articulated. In these studies, monitoring is considered a process of reviewing and editing that occurs during articulation (e.g. Butterworth, 1980). Wendel (1997) considered both planning and monitoring as activities that take place at the central processing stage of information processing and have an impact on output. He renamed them as strategic planning (i.e. planning) and on-line planning (i.e. monitoring). Strategic planning is carried out before language production while on-line planning occurs during actual production. In a sense, Wendel’s view of planning is similar to Levelt’s in that planning that takes place prior to production is differentiated from the planning (i.e. monitoring) that takes place during production.

In written production, unlike speech production, planning is not distinguished from monitoring and has been used as a more general term that covers both preparation for language use and the action of task completion. Planning is defined by Hayes & Nash (1996) as “the combination of the goal and the sequence of steps to achieve that goal” (Hayes & Nash, 1996, p. 32). It consists of two steps. The first step includes representing

the desired outcome or goal of the task and checking the resources available for carrying out the task. The second step is carrying out the task, during which alternate goals and methods are tried out until a satisfactory match between the goal and means for realizing it is found. The relationship between planning and an action of task completion is that planning guides action but does not control it. It provides suggestions for action and modifies them as action proceeds.

In writing, a distinction is made between process planning and text planning (Hayes & Nash, 1996). Process planning refers to how the writer intends to carry out the writing task (i.e. identifying strategies for fulfilling the task) while text planning focuses on the text being written. In text planning, two types of planning are further distinguished: abstract text planning and language planning. Abstract text planning takes place when writers are engaging in idea generation and conceptual planning without specifying the language to be used. In language planning, writers plan for the particular words and sentences to be used. The connection between abstract text planning and language planning is so close that it might be hard to distinguish the two. Both types of planning are involved in writing but writers might employ one type of planning more than the other at different stages of writing. Both process planning and text planning are activities involved in what will be called pre-task planning, which will be introduced in the next section. Text planning will be the principal activity involved in what will be called online planning (see next section for a discussion).

In the present study, 'planning' will be used as a general term to refer to planning that takes place both before a task (i.e. pre-task planning) is carried out and during task performance (i.e. online planning).

### **3.3 Classification of planning**

In 2005, R. Ellis proposed a way to classify planning according to the timing of planning---either before a task is carried out or during the performance of a task. Planning is classified into pre-task planning and within task planning. In the following



sections, each type of planning will be described. All the studies reviewed in this chapter adopted Ellis's (R. Ellis, 2005b) classification of planning.

### 3.3.1 Pre-task planning

There are two types of pre-task planning: rehearsal and strategic planning (R. Ellis, 2005b). Rehearsal refers to task repetition. That is, learners are given opportunities to perform the task more than once and the first performance is seen as a preparation for a subsequent performance (R. Ellis, 2005b, 2009). A number of studies (Bygate, 1996, 2001; Bygate & Samuda, 2005; Gass, Mackey, Alvarez-Torres, & Fernandez-Garcia, 1999; Kawauchi, 2005; Lynch & Maclean, 2000, 2001) were conducted to investigate the effects of task repetition on language production. What these studies aimed to find out was whether task repetition had any effect on performance of the same task.

Strategic planning refers to the pre-task activity that entails learners' making a strategic plan for the task and planning the content and language necessary for task completion without actually practicing the task prior to performing it. During strategic planning learners are usually presented with the actual task and they make use of their schemata and language resources to help them prepare for the task. In a strategic planning condition, learners may rehearse the content and language for the task. However, they usually cannot rehearse the task completely. The time provided is just enough for them to make a strategic plan.

Although both *rehearsal* and *strategic planning* are included in pre-task planning, most previous studies on planning (e.g. Crookes, 1989; Foster & Skehan, 1996; Kawauchi, 2005; Mehnert, 1998; Ortega, 1999, 2005; Skehan & Foster, 1997, 2005; Wendel, 1997; Wigglesworth, 1997; Yuan & Ellis, 2003) used the term *pre-task planning* to refer to strategic planning only. The present study is not concerned with task repetition and thus will follow the previous studies and use *pre-task planning* to refer to *strategic planning*. In this thesis, the two terms will be used interchangeably.

### 3.3.2 Within-task planning

Within-task planning, as its name suggests, refers to planning activities that are carried out during task performance. It is also referred to as online planning, which is defined as the formulating process (in terms of both content and language) and the monitoring of output (Yuan & Ellis, 2003) that a speaker or writer engages in during language production. In within-task planning, two conditions are differentiated: pressured and unpressured (R. Ellis, 2005b). Under the pressured condition, learners are given a time limit to complete a task. Their on-line planning is pressured by both time and communication needs. In contrast, unpressured within-task planning condition allows learners to take their time when carrying out the task so they can carefully plan their language on-line.

Studies on online planning usually operationalize online planning by giving learners unlimited time (e.g. R. Ellis & Yuan, 2004) or longer time than other groups of participants (e.g. D. Li, 2004) for the purpose of alleviating time pressure. However, Skehan and Foster (2005) pointed out that it is only an assumption that learners will engage in online planning when they complete tasks without time pressure. The assumption cannot be proved until there is data on learners' behavior and mental operations during task performance under an online planning condition.

The present study adopts R. Ellis' (2005b) definitions of pre-task planning and online planning discussed above. However, it is important to note that pre-task planning and within-task planning should not be considered as mutually exclusive of each other (R. Ellis, 2005b) though a distinction is made between the two. The two types of planning can be manipulated by teachers/researchers and combined in various ways, e.g. 1) pre-task planning with pressured within-task planning; 2) pre-task planning with unpressured within-task planning; 3) or unpressured within-task planning or 4) pre-task planning alone, etc. Both types of planning may ease the processing load of learners and promote language performance.

It is also important to note that it is not possible to create a pre-task planning condition where online planning is precluded or, vice versa an online planning condition where pre-task planning is precluded. Even when learners are performing under time pressure they might still engage in some online planning and when learners are carrying out a task without time pressure they could spend a little time to do pre-task planning. This is especially the case in the context of writing. Therefore, the two types of planning should be regarded as relative terms. That is, a pre-task planning condition provides relatively more opportunities for learners to engage in pre-task planning and an online planning provides relatively more opportunities to engage in online planning. The present study investigates both pre-task strategic planning and online planning in relative terms.

### **3.4 Measures for task-based language performance**

In most planning studies, the effects of task planning on language performance have been measured in terms of *complexity*, *accuracy*, and *fluency* (CAF) because they are considered as important indices that reflect the underlying interlanguage system of learners (see Skehan's notion of IL development discussed in Chapter 2). *Complexity* is defined as "the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively. This may also involve a greater willingness to take risks, and use fewer controlled language subsystems. This area is also taken to correlate with a greater likelihood of restructuring, that is, change and development in the interlanguage system" (Skehan & Foster, 1999, pp. 96-97). *Accuracy* shows "the ability to avoid error in performance, possibly reflecting higher levels of control in the language, as well as a conservative orientation, that is, avoidance of challenging structures that might provoke errors" (ibid). *Fluency* refers to "the capacity to use language in real time, to emphasize meaning, possibly drawing on more lexicalized systems" rather than the rule-based system (ibid).

The above are notional definitions of the three aspects of language performance. How they were measured in previous studies varied considerably (cf. Housen & Kuiken,

2009; Housen, Kuiken, & Vedder, 2012), as will be seen in the later sections of this chapter, making comparisons across studies difficult. In general, in most studies that measured the three aspects quantitatively *accuracy* has been measured in terms of target-like use of one or more specific grammatical structure(s) (e.g. definite article *the* and the use of plurals in Crookes' study (1989), noun modifiers and articles in Ortega's study (1999)) and/or the overall accuracy of production (e.g. percentage of error-free clauses, number of errors per 100 words). *Complexity* has been measured in terms of syntactical complexity (e.g. number of clauses per AS/C-unit or T-unit, S-nodes per utterance or T-unit, or average length of c-unit/T-unit) and lexical complexity (e.g. type-token ratio). *Fluency* has been measured in terms of temporal fluency (e.g. number of syllables per minute, pausing and silence time) and repair fluency (e.g. number of self-corrections).

There are a few studies that analyzed CAF qualitatively (e.g. Lynch & Maclean, 2000; Ojima, 2006) or measured other aspects language performance (e.g. pragmatic aspects in Nemeth & Kormos, 2001). These studies will not be reviewed in detail because their findings are not comparable to those of the present study, which used a quantitative method to measure CAF.

### **3.5 The effects of task planning on oral production**

Previous studies on task planning have disproportionately focused on oral production. Most of them (except Bygate, 2001 and Gass et al. 1999) only investigated the effects on subsequent oral performance instead of the effects of L2 development. However, they are still relevant since the present study also concerns the effects of planning on L2 production. Thus, they are reviewed in the following sections.

#### **3.5.1 The effects of rehearsal on oral production**

A few studies have investigated the effects of rehearsal (e.g. Bygate, 1996, 2001; Gass et al., 1999; Lynch & Maclean, 2000, 2001) but three of them will be reviewed

because in these studies learners' performance was measured in terms of accuracy, complexity, and fluency, making these studies comparable with other planning studies. In the following paragraphs the three studies will be reviewed one by one.

Bygate (1996) studied the language produced by one learner telling a story after viewing a video extract of a Tom and Jerry cartoon on two separate occasions. The student was required to tell the story immediately after viewing the extract. Three days later, without warning she was asked to do the same task again. Analysis of the transcripts showed that the language used in the second performance was more accurate though the margin of improvement was small. The student used more grammatically complex language and her use of lexical collocation was more native-like in the second time performance. In terms of fluency, the student's self-correcting repetitions significantly increased in the second version. These findings led Bygate to suggest that repetition could provide opportunities for learners to focus on form. The first time the learner did the task, most of the attention was focused on content and when given a chance to do the task again some processing space was freed to allow focus on form.

Gass et al. (1999) carried out a study to examine if learners would be able to use more native-like language when processing load was reduced by task repetition. The subjects were put into three groups, with two experimental groups and one control group. Experimental Group 1 watched the same video extracts three times and engaged in on-line narration of the story. At Time 4 they watched a similar but different clip and narrated the story. Experimental Group 2 watched and narrated four different extracts, with clips on Time 1 and Time 4 being the same as those seen by Group 1. The control group watched videos and narrated stories only at Time 1 and Time 4. Analysis was done on the basis of overall proficiency, morphosyntax, and lexical sophistication. Results supported the prediction that performance improved over time for Group 1, who had to repeat the task three times, but the effects were not carried over to the new task on Time 4. The study suggests that task repetition can provide learners with space to process and use more sophisticated language from their internal language system. But this effect was

quite limited and not carried over to new tasks.

Bygate (2001) investigated the effects of task repetition of two types of tasks on the structure and control of L2 language in oral production. 48 participants of the study were put into three groups. Group one did altogether six oral narrative tasks and group two did six interview tasks. Group three was the control group, who performed the tasks only in the first week and the last week of the study. The study lasted for 10 weeks. In the first week, all participants did Narrative 1 and Interview 1, which provided baseline information for the study. In the following weeks, each experimental group received treatment three times on a biweekly basis. In treatment sessions 2 and 3, they repeated one task they carried out in the previous treatment session and did a new task of the same type. In week 10, all participants were given two tasks they did in week one (Narrative 1 and Interview 1) and two new tasks (Narrative 6 and Interview 6) to do. With this design, he was able to see the longer-term effect of task repetition and if the effect of task repetition could be carried over to a new task. Learners' performance was measured in terms of accuracy, complexity, and fluency. The results of the study showed that task repetition had positive effects on complexity and fluency but not on accuracy. These effects were retained over a ten-week period. This suggests task repetition may contribute to L2 learning.

All three studies produced evidence to show that task repetition had positive effects on performance of the same task. However, as shown in Gass et. al's study (1999), the effects were not carried over to a new task. Though Bygate's study (2001) suggested that the effects of task repetition could be long lasting, the evidence was not sufficient to show that task repetition would lead to L2 learning. The results of these studies have important implications for the present study since the present study concerns itself to whether the effects of planning can be transferred to a new task and to a later time. As rehearsal is one type of pre-task planning, its effect might to some degree be comparable with that of strategic planning, which is one of the foci of the present study.

### **3.5.2 The effects of strategic planning on oral production: introduction**

Strategic planning has received most of the attention in planning research as far more studies have investigated this type of planning than online planning. In the studies that investigated strategic planning, the majority examined the effects of pre-task planning on oral production rather than on written production, which will become evident in this and later sections of this review. A total of 26 studies that investigated the effects of strategic planning on oral production will be reviewed (see Appendix I for a table summarizing these studies) and a descriptive synthesis of these studies will be presented. The studies were chosen to be included in this review because 1) they all compared the performance of a strategic planning group with that of a no planning group, and 2) they used quantitative measures for the three (accuracy, complexity, and fluency) aspects of language performance. The results of these studies will be summarized in terms of accuracy, complexity, and fluency since they are also the major variables investigated in the present study.

### **3.5.3 The effects of strategic planning on oral production: accuracy**

Of the 26 studies, 15 found some effects of strategic planning on accuracy. Among them, six found increased accuracy in some measures only (e.g. Crookes, 1989; Ortega, 1999; Sangarun, 2001). For example, among five measures for accuracy, Crookes (1989) found only a significant increase in accurate use of the definite article *the*. Ortega (1999) found a significant increase in one of the two specific measures (noun modifiers) and Sangarun (2001) found beneficial effects for pre-task planning in one of the two measures for overall accuracy (i.e. percentage of error-free clauses).

Studies that used more than one type of task in the experimental condition found significant effects for pre-task planning in some tasks (e.g. Foster & Skehan, 1996; Skehan & Foster, 1997). Foster and Skehan (1996) used three types of tasks in their study: personal information, narrative, and decision-making. They found an increase in accuracy in personal information and decision-making tasks but not in the narrative task. Skehan

and Foster (1997) found a significant increase in accuracy in only the decision-making task.

Several studies that investigated the effects of strategic planning on learners of different proficiency levels (e.g. Kawauchi, 2005; Tavakoli & Skehan, 2005; Wigglesworth, 1997) showed different results. Wigglesworth (1997) and Tavakoli and Skehan (2005), for example, found increased accuracy in learners of higher proficiency. Kawauchi (2005) recruited participants with three proficiency levels: low-intermediate, high-intermediate, and advanced. The effects of planning on accuracy were found in low and high intermediate learners but not in advanced learners.

The effects of strategic planning on accuracy also differed under different planning conditions. For example, strategic planning was operationalized as guided planning (or ‘detailed’ planning) and unguided planning (‘undetailed’ planning). Foster and Skehan (1996) found speech produced under the undetailed planning condition significantly more accurate than that produced under the no planning condition. Skehan and Foster (1999), however, found teacher-led planning, which could be considered as a strong version of guided planning, had a significant impact on accuracy. Skehan and Foster (2005) investigated the effects of detailed and undetailed planning again. Contrary to the findings of Foster and Skehan (1996), they reported accuracy effects for the detailed planning condition.

In conclusion, previous studies have produced mixed results regarding the effects of strategic planning on the grammatical accuracy of speech. The effects on accuracy of oral production have been limited because the positive effects have been shown in some measures, some tasks, under certain planning conditions, and for some learners only. Differences in measures for accuracy used, task type, learners’ proficiency level, and ways of operationalizing strategic planning (e.g. guided vs. unguided planning) seem to have influenced the results for accuracy.



### **3.5.4 The effects of strategic planning on oral production: complexity**

Research shows more evidence for the positive effect of strategic planning on complexity than it does for accuracy, though the results are still mixed. Eighteen studies out of the 26 reviewed in this section (e.g. Crookes, 1989; Foster & Skehan, 1999; Guar-Tavares, 2008; Ortega, 1995b; Sangarun, 2001; Tavakoli & Skehan, 2005; Wang, 2009; Wendel, 1997) reported clear effects for syntactic complexity while eight of them reported no effect (e.g. Elder & Iwashita, 2005; Gilabert, 2007; Mehnert, 1998; Mochizuki & Ortega, 2008; Rutherford, 2001; Tajima, 2003; Wigglesworth & Elder, 2010).

Among the 18 studies that reported positive effects for pre-task planning on syntactic complexity, Yuan and Ellis (2003) found a significant impact in only one of the three measures for complexity (number of clauses per T-unit). Foster and Skehan (1996) found clear gains in complexity in the personal information and narrative tasks but not in the decision-making task. In the following year, Skehan and Foster (1997) examined the effects of strategic planning on the three types of tasks again and found significant effects in the personal information and decision-making tasks but not in the narrative task this time. This indicates that task type could exert influence on results for complexity but how this variable interacts with planning to influence syntactic complexity needs to be researched further.

Learners' proficiency level and different pre-task planning conditions were also found to influence the results for complexity. Kawauchi (2005), for example, found clear evidence of increased syntactic complexity in low-intermediate and high-intermediate level EFL learners but not in advanced learners. In a study that investigated the effects of pre-task planning under a guided versus an unguided planning condition, Skehan and Foster (2005) reported that under the guided planning condition learners produced more subordinate clauses (the complexity measure) than the unguided and no planning conditions.

A number of studies have measured lexical complexity in addition to syntactical

complexity (e.g. Crookes, 1989; Gilabert, 2007; Ortega, 1995a, 1999; Tajima, 2003; Wang, 2009), using type-token ratio to measure lexical variety (or lexical richness). Among the eight studies that measured lexical complexity, five reported significant gains (Crookes, 1989; Gilabert, 2007; Mehnert, 1998; Ortega, 1995; Tajima, 2003;) and three did not find significant effects (Ortega, 1999, Wang, 2009; Wendel, 1997;).

To conclude, results for the effects of strategic planning on syntactic complexity in oral production were mixed. Differences in measures for complexity, task variables, learners' proficiency level, and different pre-task planning conditions again seem to influence the results for complexity. The number of studies that have measured lexical complexity was relatively small and the results were inconsistent, making it hard to draw a conclusion on the effects of planning on lexical complexity. More studies are needed to incorporate measures for lexical complexity in research investigating the effects of planning on language complexity.

### **3.5.5 The effects of strategic planning on oral production: fluency**

Among the 26 studies, 21 measured temporal fluency and in 14 there was an increase in fluency as a result of planning (e.g. Foster & Skehan, 1999; Mehnert, 1998; Ortega, 1999; Tajima, 2003; Tavakoli & Skehan, 2005; Wang, 2009; Wendel, 1997; Yuan & Ellis, 2003). Ten studies measured repair fluency and two of them found a significant reduction in self-corrections (Ortega, 1995; Wigglesworth, 1997) while eight did not find such an effect (e.g. Foster & Skehan, 1996, 1999; Wang, 2009). It seems that planning has more effects on temporal fluency than on repair fluency.

Variables that have been found to influence the results for accuracy and complexity also affected the results for fluency in general. For example, the learners' proficiency level exerted an influence on results for fluency (Kawauchi, 2005; Ortega, 1999; Wigglesworth, 1997) but the findings were inconsistent. Ortega (1999), who studied the effects of planning on advanced learners, found an increase in fluency. Wigglesworth (1997) reported greater effects of planning on fluency for higher level learners than for

the lower level learners. However, Kawauchi (2005) found increased fluency in low and high intermediate level learners but not in advanced learners. It has to be noted that proficiency in previous studies has been defined according to different criteria. Some studies used a program specific course level to indicate learners' proficiency (e.g. Foster & Skehan, 1996; Ortega, 1999) while some others established proficiency level by test scores, such as TOEFL or IELTS scores (e.g. Crookes, 1989; Wigglesworth, 1997). A conclusion on the effects of planning on learners of different proficiency levels will not be possible until proficiency is defined with a uniform standard.

Task variables also influenced the results for fluency. This is most evident in the early works of Foster and Skehan (e.g. Foster & Skehan, 1996, Skehan & Foster, 1997). Increased fluency was found in the personal information and narrative tasks but not in the decision-making task. Mehnert (1998) also reported greater effects for fluency in a more structured task as opposed to a less structured tasks.

To recapitulate, the effects of planning on fluency were more evident in temporal fluency than in repair fluency. Learners' proficiency level and task type caused variability in results for fluency.

### **3.5.6 The effects of online planning on oral production**

Of the 26 studies that have been reviewed in the previous section, only three also investigated the effects of online planning on oral production and the results are inconsistent. Yuan and Ellis's study (2003) found that on-line planning enhanced accuracy while Wang (2009) and Nakakubo (2011) did not find such an effect. None of the studies found any effect of online planning for complexity or fluency.

It is noteworthy that online planning was operationalized differently in these studies. Yuan and Ellis (2003) and Nakakubo (2011) allowed unlimited time for the online planners to carry out the tasks (unpressured online planning) whereas Wang (2009) investigated the effects of time-pressured online planning by asking the participants in the online planning group to tell a story as they were watching a video that was edited to play

at a slower rate. Moreover, Nakakubo (2011) provided a practice session for the online planners, in which the online planners were prompted to think about what they were going to say every 30 seconds, while Yuan and Ellis (2003) and Wang (2009) did not provide such a practice. Nakakubo's (2011) study also found that the time pressure variable influenced the results of the study, the fluency aspect in particular. This indicates that time-pressured online planning could have different effects on oral performance than unpressured online planning. This should be taken into consideration in future research.

### **3.5.7 The process of planning: how do learners plan**

To better understand why planning has or does not have effects on task performance, data on how learners planned was collected in some studies. Nine studies (Guará-Tavares, 2008; Kawauchi, 2005; Nakakubo, 2011; Ortega, 1999, 2005; Sangarun, 2001; Wendel, 1997; Wigglesworth & Elder, 2010; Yuan & Ellis, 2003) out of the 26 studies investigated what learners did during pre-task planning. It has been found that in general learners plan both content and form during planning but their attention is not equally divided. They tend to engage in content planning more than language planning and in language planning more attention is allocated to selecting lexical items. A general pattern of planning was also found (Ortega, 1999). That is, learners tend to plan the main ideas first, and then engage in planning the organization of the output, and last plan the details. These findings provide insight into why and how planning has (or not have) effects on oral production. For example, Yuan and Ellis (2003) argued that content planning resulted in enhanced complexity in their study. Ortega (1999) attributed increased fluency and complexity in her study to the enhanced attention to form brought about by pre-task planning. However, these studies did not investigate how planning affected the subsequent the production process, causing a limit on our understanding of planning. The present study goes some way towards addressing this issue.

### 3.5.8 Summary

Studies investigating the effects of planning on oral production have been conducted in both foreign language teaching (FLT) and second language teaching (SLT) contexts and the number of studies conducted in these contexts was relatively balanced (12 SLT and 14 FLT). Most of these studies used oral narrative tasks to elicit oral production (15 out of 26) and most of them operationalized pre-task planning by giving participants 10 minutes to prepare prior to the tasks (14 out of 26)<sup>5</sup>. Online planning has been operationalized as giving participants in the online planning group unlimited time or more time than other groups to complete the tasks. Almost all of the studies (except Skehan & Foster, 2005<sup>6</sup>) only investigated the immediate effects of planning on oral performance without concerning the durability of the effects. This is a serious limitation in that without knowing whether the effects of planning can be sustained for a longer time and whether the effects can be transferred to a new context, the argument for a facilitative role for planning in L2 development still remains a hypothesis. Another limitation of previous studies is that most of them (except the few discussed in the previous section) only looked at the products by comparing planned speech with unplanned speech rather than adopting a process-product approach, which involves an investigation of the planning process and the process of subsequent production. It is important to study these processes because data on these processes can help us better understand results of studies.

In terms of the effects of planning on oral performance, pre-task planning has benefited fluency and complexity more than accuracy and there is a little evidence to suggest that online planning benefits accuracy rather than complexity or fluency.

While studies on the effects of planning on oral production have provided insights into how humans process language during speech production, the findings cannot be

---

<sup>5</sup>See Appendix I for how much time was given for pre-task planning in other studies.

<sup>6</sup>Skehan and Foster's study (2005) compared learners' speech in the first five minutes with that in the next five minutes and found that learners' performance in the second five minutes deteriorated markedly, suggesting that the effects of planning could not be retained for a long time.

generalized to the context of written production. Though it has been assumed that processes involved in oral production and written production have much in common (R. Ellis & Yuan, 2004; Kellogg, 1996), there are important differences between the two modes of production. As has been discussed in Chapter 2, there are a number of differences between speaking and writing, and one crucial difference is that speaking has to be completed in real time, whereas writing allows more time on the part of the writer to control formulation and monitoring. In other words, a writer is under less communicative pressure than a speaker and has more control over time to engage in strategic planning or online planning during production. Therefore, the effects of planning on written production should not be expected to be the same as the effects on oral production.

### **3.6 Impact of planning on L1 writing**

As has been explained in section 3.2, unlike in speech production, planning in writing is considered as an important strategy that could occur at any stage of writing (Flower & Hayes, 1980; Hayes & Nash, 1996; Kellogg, 1996). A writer may plan his/her text before starting to compose and continue to plan online as s/he writes. It is commonly recognized that writers engage in both pre-task planning and online planning for writing tasks.

In general, three planning activities have been identified as sub-processes of planning: idea generation, organization, and goal-setting (Flower & Hayes, 1980). Idea generation, though it can occur prior to composing, has been investigated as a within-task (online planning) activity with think-aloud protocol being an elicitation technique. It was found that generating ideas in the early phase of composing (i.e. first one third of composing time) positively correlated with the quality of texts written by adolescent L1 writers while generating ideas in the late stage of composing influences text quality negatively (van der Hoeven, 1999). Moreover, it was found that generating multiple, related ideas as opposed to simple idea generation has a positive impact on teenagers' L1 writing quality

(Rijlaarsdam & van den Bergh, 2006; van den Bergh & Rijlaarsdam, 1999). The results of these studies suggest that devoting time and attention to content preparation at the early phase of or prior to writing would be beneficial to subsequent writing.

While idea generation has been studied as a within-task planning activity, organizational techniques, such as outlining and clustering of ideas, have been examined as pre-task planning activities. A series of studies were conducted by Kellogg (1987, 1988, 1990) to investigate the effects of various types of pre-writing activities on text quality and writing fluency. The strategies investigated in the 1987 and 1988 studies were outlining and composing a rough draft during the prewriting period. The hypothesis was that these two were cognitive strategies that could lessen a writer's workload, thus leading to improved writing performance. In his 1987 study, eighteen college students were randomly assigned to one of four strategy groups: no outline versus outline, rough draft versus polished draft. They were asked to engage in the assigned prewriting activities in a letter writing task. The no outline group started writing after receiving instructions and the outline group had 5-10 minutes to prepare a written outline before writing the letter. The rough draft group was instructed to compose a rough draft first with the chief aim of getting their thoughts on paper rather than being concerned with how well their ideas were expressed. The polished draft group was directed to produce a polished draft with the chief aim of expressing their ideas as well as possible from the beginning. The student's writing process, efficiency (i.e. the time needed to produce a document), and the quality of their written texts were examined. Efficiency was measured by composing time and words per minute. The quality of written texts was measured by using a holistic scoring method. Two readers were asked to make judgments on five dimensions of the texts (i.e. language usage, organization coherence, idea development, effectiveness of communication, and mechanics). The results showed that preparing a written outline, compared with the no outline condition, increased the time students spent on translating ideas into text, improved the quality of the written texts but failed to increase the efficiency of their writing. In the rough draft vs. the polished draft conditions, participants differed in how they composed

the texts. However, these processing differences had no influence on text quality or efficiency.

Kellogg's 1988 study further examined whether and how preparing a written outline and composing a rough draft during prewriting eased attentional overload and consequently enhanced writing performance. To achieve this aim, two experiments were conducted. In experiment 1 the four task conditions were the same as the ones in his 1987 study: outline vs. no outline, rough draft vs. polished draft. The processing time and effort given to planning, translating ideas into text, and reviewing ideas and text were recorded using directed retrospective report. Results of this experiment showed that the outline strategy improved the quality of written products measured by holistic rating but did not affect the efficiency of writing (i.e. writing fluency). Composing a rough draft did not influence the quality of written products or the efficiency. This suggests that in L1 writing, as a pre-writing strategy outlining, which involves less linguistic encoding of ideas, is more effective than writing a rough draft, which involves more linguistic encoding.

Experiment 2 aimed to examine the cognitive effort the students expended when writing under different planning conditions, as well as replicating the effects of outlining on writing. In this experiment, students wrote a persuasive essay. Three task conditions were tested: no outline, written outline and mental outline. In the no-outline condition, students began writing as soon as they received instruction. In the mental outline condition, students were given pre-writing planning time as the written outline group, but they were not allowed to write anything down. To record the cognitive effort students expended during writing, the students were asked to report their thoughts occasionally on hearing a 'beep' sound generated by a computer. They then needed to press one of four labeled buttons to indicate which process of writing (i.e. planning, translating ideas into text, reviewing, or others) they were engaging in at that moment. The results showed that outlining did not influence the efficiency of writing. The written and mental outline conditions did not yield any differences. Overall, students spent more effort on planning



than translating or reviewing. Students in both outline conditions spent more time on translating ideas into text than those in the no planning condition. This finding indicates that outlining can free up some space in working memory so students can focus more on translating ideas into text during the subsequent production.

In 1990, Kellogg (1990) compared the effects of two pre-writing activities, outlining and clustering, on the quality of essays of college students. 207 college students were divided randomly into nine groups and asked to write an informative essay. One group (control) started drafting without any pre-writing time, the second group had time to prepare an outline before drafting, and the third group did clustering of ideas before they wrote. Task demand was also a variable under investigation. Three task conditions varying in the degree of demands they placed on the learners were created. In the first condition students were provided with only the topic of the essay. In the second condition, students were provided with possible ideas for inclusion as well as the topic. In the third condition, students were given a suggested organizational pattern in addition to the topic and possible ideas. It was assumed that the task demand would be the highest in condition one and the lowest in condition three. Students' essays were judged by two independent raters holistically in terms of content and style. The results showed that participants in the outline condition wrote higher quality essays in terms of content than did those in the cluster and the control conditions. The gain in quality through outlining decreased with the decrease of task demands. That is, the beneficial effect of outlining was strongest in the topic only condition; it decreased as the task demands were lessened by provision of ideas; and the effect became nonexistent for the topic plus ideas plus organization condition. In terms of fluency, the outline group also wrote longer essays than the control and cluster groups. However, when fluency was measured by words per minute with pre-writing time included in the calculation, the significant effect of outlining disappeared. The study confirms the beneficial effects of prewriting planning on text quality and some effects on fluency of writing. It also suggests that task demand would affect the effects of planning.

Kellogg's finding was supported by Rau and Sebrechts (1996), who found an indirect

impact of outlining on the writing of college students. Rau and Sebretchts (1996) conducted a two-part experiment to investigate the effects of pre-task planning on university level L1 writing. The first experiment is irrelevant to the present study. In their second experiment, they compared texts composed under the outlining conditions and those composed under a no planning condition. They also compared the mental operations of the participants during the composing process. Half of the participants were asked to do think-aloud while performing the tasks. The think-aloud protocols were analyzed for statements related to content generation and conceptual planning (e.g. structuring and ordering of ideas, giving directions to oneself for completing the task, identifying audience beliefs and effectiveness of writing). Results of analysis revealed that participants in the outlining condition produced more statements related to conceptual planning than those who did not use an outline. Participants' essays were assessed in terms of the overall quality of the essays and performance in the areas of creativity, technical concepts, grammar and clarity by three independent raters. A multiple regression analysis revealed that the percentage of conceptual planning statements was a significant predictor of overall quality ratings. Given the positive correlations between outlining, the percentage of conceptual planning statements, and the quality of essays, these results indicated that outlining benefited L1 writing quality by providing writers with the opportunity to engage in conceptual planning, in which they could organize their ideas and evaluate audience's needs and effectiveness of writing. Moreover, with time to do pre-task planning (i.e. outlining) the participants made less revision in content during writing.

Kellogg's studies (1987, 1988, 1990) and Rau and Sebretchts 's study (1996) all showed positive effects of outlining as a pre-task planning strategy on the overall quality of L1 writing. Outlining also had some effects on writing fluency as measured by the speed of writing. These results justify the hypothesis that pre-task planning may have positive effects on L2 writing because the processes of L1 writing and L2 writing are similar in general (e.g. Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; Kellogg, 1996; Zimmerman, 2000). This is supported by studies that compare planning behaviors of L1

and L2 writers.

### **3.7 Comparison of planning in L1 and L2 writing**

Studies have suggested that L1 and L2 writing planning are similar in terms of allocation of planning time during composition (Jones & Tetroe, 1987; Manchon & Roca de Larios, 2007; Skibniewski & Skibniewska, 1986; Whalen & Menard, 1995) and planning activities (e.g. idea generation and goal setting) taking place in different phases of writing (Manchon & Roca de Larios, 2007). Manchon and Roca de Larios (2007), for example, found that planning in both L1 and L2 writing took place mainly in the first one third of writing time.

While the studies mentioned above found quantitative similarities between L1 and L2 writing planning, some other studies found qualitative differences (Chenoweth & Hayes, 2001; Whalen & Menard, 1995). Whalen and Menard (1995) compared planning in L1 and L2 writing at three levels of discourse: 1) pragmatic level, in which the rhetorical and contextual parameters of the text were planned, 2) textual level, in which information that made up elements including textual macrostructure and cohesion and coherence of text was processed, and 3) linguistic level, in which writers retrieve, formulate, and manipulate morphosyntactic and lexical linguistic elements. The study found that pragmatic planning occurred twice as frequently in L1 writing as in L2 writing whereas linguistic planning took place three times as frequently in L2 writing as in L1 writing. That is, when writing in L1 writers were able to engage in more global processing of information while they primarily engaged in planning at a linguistic level when writing in L2.

### **3.8 Impact of planning on L2 writing**

Compared with the number of studies that have investigated the effects of task planning on oral production, the number of studies that have examined the effects of task planning on written production is much smaller. Twelve studies have been found to look into the effects of planning on written production (Al-Humaidi, 2008; R. Ellis & Yuan,

2004; Farahani & Meraji, 2011; M. D. Johnson, 2011; D. Li, 2004; Z. Li, 2008; Ojima, 2006; Ong & Zhang, 2010; Pu, 2009; Rahimpour & Safarie, 2011; Shi, 1998; Shin, 2008) in an L2 context (see Table 1 for a summary of these studies). Two of them (Shi, 1998; Shin, 2008) studied the impact of group planning on the quality of writing. The quality of written texts was measured by impressionistic ratings. Shin (2008) found that collaborative planning led to higher holistic scores in expository writing than individual planning but such an effect was not found in argumentative writing. Shi (1998) found that pre-writing discussion provided more perspectives and more ideas for writers. Since the present study concerns individual planning and the participants' writing performance was measured in terms of accuracy, complexity, and fluency as in most of the other previous studies, these two studies will not be reviewed in more detail. Li's study (2008) used a questionnaire to collect information on learners' metacognitive knowledge about planning, their planning strategies, and planning time. Then the causal relations between these variables and the text quality measured with holistic scoring were established. There was no comparison between a planning condition and a non-planning condition and the texts were not analyzed in terms of accuracy, complexity, and fluency. For this reason, Li's study is not going to be reviewed in more detail either.

The nine studies, which will be reviewed in this section, all had at least a pre-task individual planning condition compared with another planning condition (pre-task planning vs. no planning or pre-task planning vs. online planning) and the writing performance was measured in terms of two or all three aspects of language performance (i.e. accuracy, complexity, and fluency). Among the nine studies, four also investigated an online planning condition. Table 1 summarizes the nine studies. The impacts of pre-task planning and online planning will again be reviewed separately.

### **3.8.1 Impact of pre-task planning on L2 writing: accuracy**

As in the studies on the effects of planning on oral production, accuracy has been measured in terms of overall accuracy of texts (e.g. number of errors per 100 words,

percentage of error-free clauses) and in some studies the use of some specific grammatical feature(s) (i.e. English articles). Of the nine studies, two did not measure the effects of pre-task planning on accuracy (M. D. Johnson, 2011; Ong & Zhang, 2010). In the other seven studies, only one study (Farahani & Meraji, 2011) found a significant positive effect of pre-task planning on accuracy in the global measures but not in the specific measure of accuracy (i.e. English articles). Increased accuracy was not found in all other eight studies, indicating that pre-task planning did not promote L2 writing accuracy.

### **3.8.2 Impact of pre-task planning on L2 writing: complexity**

As has been pointed out in the previous section, complexity has been measured in terms of syntactic complexity and lexical complexity (see Table 1 for measures). One study (Ong & Zhang, 2010) measured only lexical complexity. For the other eight studies, three reported increased syntactic complexity under the pre-task planning condition (R. Ellis & Yuan, 2004; Farahani & Meraji, 2011; Ojima, 2006) while the other five reported no difference between the pre-task planning group(s) and the non-pre-task planning group(s) (Al-Humaidi, 2008; M. D. Johnson, 2011; D. Li, 2004; Pu, 2009; Rahimpour & Safarie, 2011).

In terms of lexical complexity measured by lexical variety, of the six studies that measured this aspect of language performance (Al-Humaidi, 2008; Ellis & Yuan, 2004; Farahani & Meraji, 2011; M. D. Johnson, 2011; D. Li, 2004; Ong & Zhang, 2010) only one study (Ong & Zhang, 2010) found significant differences between groups. Ong and Zhang (2010) compared writing performance with pre-task planning time and without pre-task planning time. In their study, there were four task conditions: extended pre-writing planning (20 minutes to plan and 10 minutes to write), pre-writing planning (10 minutes to plan and 20 minutes to write), free writing (no pre-writing planning time and 30 minutes to write), and a control condition (30 minutes to write). The difference between the free writing condition and the control condition was that in the free writing

condition participants were instructed to write immediately and continuously for 30 minutes whereas the participants in the control group were instructed to finish the first draft in their usual style (whether to spend a little composing time on pre-task planning and whether to write continuously were up to the participants) in 30 minutes. This study found that the participants in the free writing condition outperformed those in the two pre-task planning conditions in lexical variety and the control condition outperformed the extended pre-task planning condition in this aspect as well. It should be noted that the two groups with pre-task planning time were allowed shorter time to complete the tasks, which means that they carried out the task under time pressure and they had less opportunity to do online planning. The contrast between groups could be understood as a contrast between pre-task planning and online planning. Because of this research design, the effects of pre-task planning of this study are not comparable with those of other studies, in which the pre-task planning group had the same amount of time to complete the tasks as the no planning group. Except for this study, no significant between-group differences in lexical complexity were reported in the studies that investigate the effects of pre-task planning on L2 writing.

In sum, pre-task planning does not seem to have much effect on L2 writing complexity since the majority of studies mentioned above did not report significant increase in complexity in the pre-task planning condition. The number of studies is small. More studies on the effects of planning on L2 written texts are needed to verify this finding.

### **3.8.3 Impact of pre-task planning on L2 writing: fluency**

In planning studies, fluency has been measured in terms of temporal fluency (i.e. rate of production) and repair fluency (i.e. dysfluency). Seen from the results for these measures, the effects of pre-task planning on L2 writing fluency were more evident than those on accuracy or complexity. Six out of the nine studies reported positive effects of pre-task planning on temporal fluency (Al-Humaidi, 2008; Ellis & Yuan, 2004; Farahani

& Meraji, 2011; M. D. Johnson, 2011; Ojima, 2006; Rahimpour & Safarie, 2011) and one study (Ong & Zhang, 2010) found the effect in the opposite direction. Ong and Zhang's study (2010) found that the free writing condition led to higher writing speed than the two pre-task planning conditions and the control condition also benefited writing speed. They explained the results in terms of task complexity in relation to Robinson's resource-dispersing dimension of task complexity. They argued that the availability of planning time increased task complexity instead of decreasing it. Therefore, contrary to their hypothesis, the free writing condition turned out to be the least complex task and the extended pre-task planning was the most complex task condition. Their operationalization of reducing task complexity by providing pre-task planning time was problematic. Providing pre-task planning time would decrease task complexity (Robinson, 2001, 2003, 2005) only when within-task planning opportunity is held constant. Otherwise, the potential effects of pre-task planning would be diminished. That is, if the pre-task planners had more time pressure than the non-planners, the advantage provided by planning opportunity might be mediated by the time pressure. Time pressure could be a variable that also influences task complexity.

With respect to dysfluency, three studies (Al-Humaidi, 2008; R. Ellis & Yuan, 2004; Farahani & Meraji, 2011) measured this aspect of performance. Two of the three (R. Ellis & Yuan, 2004; Farahani & Meraji, 2011) reported reduced dysfluency as a result of pre-task planning.

To conclude, pre-task planning seemed to have some effect on fluency of writing. To summarize the overall effects for pre-task planning on L2 writing performance, it did not have much benefit for accuracy and had minimum effects on complexity, while having more effects on fluency of writing.

### **3.8.4 Impact of online planning on L2 writing**

As has been explained earlier, four studies also examined the effects of online planning on L2 writing performance (Al-Humaidi, 2008; R. Ellis & Yuan, 2004; D. Li,

2004; Rahimpour & Safarie, 2011) in the nine studies reviewed in the previous section. Among them, only Ellis and Yuan's study (2004) found that online planning promoted accuracy. In the three groups of participants (i.e. pre-task planning, online planning, and no planning), the online planning group had the highest ratio of error-free clauses. Though the between-group difference was not statistically significant, the effect size was large. The online planning group also significantly outperformed the no planning group in the use of correct verb forms. Based on these results, they concluded that online planning had beneficial effects on accuracy.

In terms of complexity, no effect of online planning was found. None of the four studies reported any significant increase in complexity.

With respect to fluency, three studies (Al-Humaidi, 2008; R. Ellis & Yuan, 2004; Rahimpour & Safarie, 2011) found an effect for online planning. Ellis and Yuan (2004) found that online planning had positive effects on dysfluency. The online planning group had a significantly lower ratio of self-corrections than the no planning group. However, online planning reduced temporal fluency. The online planning group was significantly slower than the pre-task planning group while the difference between this group and the no planning group was not significant. Al-Humaidi's (2008) study found a significant between-group difference in temporal fluency with the pre-task planning and the no planning group outperforming the online planning group. Rahimpour and Safarie's study (2011) also found the pre-task planning group significantly outperformed the online planning group in temporal fluency. Li's study (2004) was the only one that did not find a significant difference between the pre-task planning and the online planning group.

An important difference between Li's study (2004) and Ellis and Yuan's study (2004) was the operationalization of the online planning condition. In Ellis and Yuan's study (2004) the online planning group was given unlimited time to write whereas Li (2004) gave participants in the online group more time than the pre-task planning group to write but with a time limit. The rationale for making this change was to limit the opportunity for pre-task planning on the part of the online planners so that the two planning



conditions could be differentiated.

In conclusion, online planning did not seem to have much effect on accuracy or complexity of L2 written production. It decreased the temporal fluency of writing.

Table 1 Studies that investigated the effects of planning on L2 written production

Study	Learners and contexts	Time span of the main study	Research questions or hypotheses	Method and instruments	Measures	Results (quantitative)
Ellis and Yuan (2004)* <sup>7</sup>	42 Chinese learners of English at a Chinese university/EFL	Less than 1 day (pre-test + one writing session of less than 30 minutes)	What are the effects of pre-task planning and online planning on L2 learners' production of written narratives?	<ul style="list-style-type: none"> <li>➤ Between groups design.</li> <li>➤ Three planning conditions: no planning (NP), pre-task planning (PTP), online planning (OP)</li> <li>➤ Planning time: 10 minutes for pre-task planning; unlimited for online planning.</li> <li>➤ Task: a written narrative task</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Syllables per minute;</li> <li>● Ratio of dysfluencies</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses /T-unit; (C/T)</li> <li>● Number of different verb forms; (VF)</li> <li>● Mean segmental type-token ratio (MSTTR)</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses;</li> <li>● Percentage of accurately used verb forms</li> </ul>	<p>1. No significant differences in error-free clauses. In correct verb forms, significant differences were found: OP&gt;NP</p> <p>2. No significant difference in C/T. In VF, PTP significantly outperformed NP.</p> <p>No significant difference in lexical variety (MSTTR).</p> <p>3. PTP wrote significantly faster than OP and NP. PTP and OP's dysfluency rates were lower than NP's.</p>
Li (2004)*	24 learners of English at an ESL program in New Zealand/ESL	Same as Ellis and Yuan's study (2004)	What are the effects of pre-task planning and online planning on L2 writing performance?	<ul style="list-style-type: none"> <li>➤ Between groups design.</li> <li>➤ Two planning conditions: pre-task planning; online planning</li> </ul>	Same as Ellis and Yuan's study (2004)	No significant difference was found in accuracy, complexity, or fluency between the two groups.

<sup>7</sup>Studies that are marked with \* investigated the effects of both pre-task planning and online planning.

				<ul style="list-style-type: none"> <li>➤ Planning time: 10 minutes for pre-task planning; 30 minutes for online planning</li> <li>➤ Task: a written narrative task.</li> </ul>		
Ojima (2006)	3 Japanese learners of English at an ESL school in Toronto/ESL	4 weeks (three in-class writing sessions of less than 15 minutes + one 15-30 minute session completed at home)	To what extent does concept-mapping affect L2 writing performance?	<ul style="list-style-type: none"> <li>➤ Case study.</li> <li>➤ Two planned and two unplanned writing pieces were collected from each person.</li> <li>➤ Planning time: 10 minutes for one planned essay; the other was a take-home assignment</li> <li>➤ Task: task type not reported.</li> <li>➤ The concept maps were also collected.</li> </ul>	<p>Holistic ratings</p> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of words;</li> <li>● Word variation;</li> <li>● Type-token ratio (TTR)</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Words p/T-unit;</li> <li>● Clauses /T-unit</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses;</li> <li>● Percentage of grammatical error-free clauses;</li> <li>● Percentage of lexical error-free clauses</li> </ul>	<ol style="list-style-type: none"> <li>1. In holistic scores, 2/3 students got higher scores for planned output.</li> <li>2. No difference in accuracy was found.</li> <li>3. Planned compositions were better in complexity.</li> <li>4. Planned compositions were more fluent.</li> </ol>
Al-Humaidi (2008)*	30 Saudian learners of English at a university in Saudi	2 weeks (one pre-test + one writing session of	What are the effects of strategic planning and online planning on written narrative task performance?	<ul style="list-style-type: none"> <li>➤ Between groups design</li> <li>➤ Planning conditions: no planning; Strategic planning;</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of syllables/minute</li> <li>● Number of dysfluencies</li> </ul> <p>Complexity:</p>	<ol style="list-style-type: none"> <li>1. No effects of either pre-task planning or online planning for accuracy were found.</li> <li>2. No effects of either</li> </ol>

	Arabia/EFL	less than 30 minutes)		<p>online planning</p> <ul style="list-style-type: none"> <li>➤ Planning time: 10 minutes strategic planning; unlimited for online planning</li> <li>➤ Task: a written narrative</li> </ul>	<ul style="list-style-type: none"> <li>● Clauses/T-unit</li> <li>● Verb forms/total number of verbs</li> <li>● Mean segmental type-token ratio</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses</li> <li>● Percentage of accurately used verbs</li> </ul>	<p>pre-task planning or online planning for complexity were found.</p> <p>3. The pre-task planning group and the no planning group wrote significantly faster than the online planning group. No significant difference was found for dysfluency.</p>
Pu (2006)	24 Chinese learners of English in a Chinese university/EFL	Less than 1 day (one writing session of about 50minutes)	How does pre-writing discussion in L1 and L2 affect language quality of argumentative compositions produced by Chinese English majors?	<ul style="list-style-type: none"> <li>➤ Between groups design</li> <li>➤ Four planning conditions: discussion in English; discussion in Chinese; discussion in English/Chinese; no discussion</li> <li>➤ Pre-writing discussion time: 15 minutes</li> <li>➤ Task: an argumentative writing task</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Total number of words</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Number of clauses/T-unit</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Number of errors per 100 words</li> </ul>	No significant difference was found in accuracy, complexity, or fluency between groups.

<p>Ong, J. &amp; Zhang (2010)</p>	<p>108 Chinese learners of English at a university in Singapore/EFL</p>	<p>Less than 1 day (one writing session of about 55 minutes)</p>	<p>What are the effects of task complexity manipulated along dimensions of +/-planning time, +/-ideas and macro-structure and +/-draft availability on fluency and lexical complexity in EFL students' argumentative writing?</p>	<ul style="list-style-type: none"> <li>➤ Between groups design</li> <li>➤ Planning conditions: Extended pre-task planning (20 minutes planning + 10 minutes writing); Pre-task planning (10 minutes plan +20 minutes writing); Free-writing (30 minutes writing); Control (30 minutes writing)</li> <li>➤ Provision of the ideas and macro-structure factor: Topic, ideas, and macro-structure given; Topic and ideas given; Topic given;</li> <li>➤ Draft availability factor: Draft available vs. draft unavailable</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Fluency I: mean number of words produced per minute of transcription.</li> <li>● Fluency II: mean number of words produced per minute out of the total time spent on task.</li> </ul> <p>Lexical complexity:</p> <ul style="list-style-type: none"> <li>● Word types squared divided by the total number of words.</li> </ul>	<ol style="list-style-type: none"> <li>1. The planning conditions led to significantly lower scores in fluency II and lexical complexity than the no planning conditions (i.e. free writing and the control condition);</li> <li>2. Provision of ideas and macro-structure, had no significant effect on fluency or lexical complexity.</li> <li>3. Decreasing task complexity through draft availability produced no significant differences in fluency or lexical complexity.</li> </ol>
-----------------------------------	---	--	---	--	--	--

				➤ Task: An argumentative writing task		
Johnson (2011)	1084 Spanish-speaking learners of English at an English language institute in Peru/EFL.	2 weeks (one baseline data collecting session of about 45 minutes + one experimental session of about 55 minutes)	What are the effects of three pre-writing strategies on L2 writing texts?	<ul style="list-style-type: none"> <li>➤ Between groups design</li> <li>➤ Five conditions: No planning (control); Idea generation; Organization (complete an outline); Goal setting; Goal setting +organization</li> <li>➤ Pre-task planning time: 10 minutes planning</li> <li>➤ Task: Two argumentative writing tasks</li> </ul>	<p>Holistic scoring</p> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Total number of words</li> <li>● Average length of sentences</li> </ul> <p>Lexical sophistication:</p> <ul style="list-style-type: none"> <li>● Lexical diversity (using MTLTD computer program)</li> <li>● Type-token ratio of content words</li> <li>● Ratio of pronouns to noun phrases (using Coh-Metrix computer program)</li> </ul> <p>Grammatical complexity:</p> <ul style="list-style-type: none"> <li>● Noun phrases per 1,000 words</li> <li>● Words before the main verb of clauses</li> <li>● Flesch reading ease</li> </ul>	<ol style="list-style-type: none"> <li>1. The organization and idea generation pre-task planning conditions had minimal impact on grammatical complexity.</li> <li>2. The organization condition had a significant effect on fluency.</li> <li>3. Pre-task planning had no impact on holistic scores or lexical sophistication.</li> </ol>
Rahimpour and Safarie (2011) *	37 Iranian learners of English at a university in Iran/EFL	Less than 1 day (one writing session of about 30 minutes)	What are the effects of pre-task planning and online planning on descriptive writing of EFL?	<ul style="list-style-type: none"> <li>➤ Between groups design</li> <li>➤ Two planning conditions: pre-task planning; online</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of T-units per text;</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Percentage of dependent clauses</li> </ul>	<ol style="list-style-type: none"> <li>1. No significant difference between the two groups in accuracy.</li> <li>2. No significant difference between the two groups in</li> </ol>

		minutes)	learners?	<p>planning</p> <ul style="list-style-type: none"> <li>➤ Planning time: 10 minutes for pre-task planning; unlimited for online planning</li> <li>➤ Task: a descriptive writing task</li> </ul>	<p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free T-units</li> </ul>	<p>complexity.</p> <p>3. The pre-task planning group outperformed the online planning group in fluency.</p>
Farahani and Meraji (2011)	123 Iranian learners of English at a language institute in Iran/EFL	Less than 1 day (one writing session of about 24 minutes)	What are the effects of manipulating task complexity along the dimension of +/- planning and +/- Here-and-Now on L2 writing performance?	<ul style="list-style-type: none"> <li>➤ Between groups design</li> <li>➤ Four task conditions: + Here-and-Now &amp; + planning; + Here-and-Now &amp; +-planning; -Here-and-Now &amp; + planning; -Here-and-Now &amp; - planning</li> <li>➤ Planning time: 10 minutes</li> <li>➤ Task: a narrative writing task</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Length of text</li> <li>● Number of dysfluencies</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● S-nodes/T-unit (S-nodes/T)</li> <li>● Clauses /T-unit (C/T)</li> <li>● Percentage of lexical to function words (%L/F)</li> <li>● Mean segmental type-token ratio (MSTTR)</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses (EFC)</li> <li>● Number of errors per 100 words (NER)</li> <li>● Target-like use (TLU) of English articles</li> </ul>	<p>1. Planning significantly enhanced accuracy in EFC and NER. The TLU was not significantly different across groups.</p> <p>2. Planning significantly promoted syntactic complexity (S-nodes/T and C/T). No significant differences were found for lexical complexity (%L/F and MSTTR).</p> <p>3. Planning significantly reduced the number of dysfluencies and encouraged longer texts.</p> <p>4. The +/- Here-and-Now variable did not have a significant effect on accuracy or complexity. It had a</p>

						significant effect on fluency with better performance found in the +Here-and-Now task.
--	--	--	--	--	--	--



### 3.8.5 Summary

The table above summarizes studies investigating the effects of task planning on L2 writing. It can be seen that these studies were predominantly conducted in EFL contexts (EFL:ESL=7:2). The tasks used in the studies were written narratives (Al-Humaidi, 2008; R. Ellis & Yuan, 2004; Farahani & Meraji, 2011; D. Li, 2004), argumentations (M. D. Johnson, 2011; Ong & Zhang, 2010; Pu, 2009), or descriptive writing (Rahimpour & Safarie, 2011).

In most of these studies (seven out of nine), pre-task planning was operationalized as giving 10 minutes for the participants to prepare for the subsequent writing and online planning was operationalized as giving participants unlimited time to write in three out of four studies (except D. Li, 2004). The operationalizations of planning have implications for the present study, which will be elaborated in the following section.

Like the studies that investigated the effects of task planning on oral production, all planning studies conducted in L2 writing contexts only investigated the immediate effects of planning by looking at participants' performance in one writing session (except Ojima, 2006). This is a serious limitation because with this design, how learners perform in a new task or at a later time cannot be examined. The issue of the effects of planning on L2 writing development is untackled.

Another limitation planning studies on L2 writing share with planning studies on oral production is that the process of planning and writing was not examined sufficiently, for only two studies (Al-Humaidi, 2008; R. Ellis & Yuan, 2004) out of the nine reviewed collected data on what learners were doing when they planned.

With respect to the effects of planning on L2 writing performance, pre-task planning has limited effect on accuracy and complexity. It has some positive effects on fluency, temporal fluency especially. Online planning does not have much effect on accuracy and complexity and it has a negative effect on temporal fluency.

In conclusion, unlike the studies on planning in L1 writing, studies on planning in L2 writing found little positive effect on the quality of writing measured in terms of accuracy

or complexity. It should be noted, however, that the number of studies on planning in L2 writing is still quite small. The evidence so far is not sufficient for a conclusion to be drawn confidently. More studies are needed in this line of research.

### **3.9 Summary of review**

1) Most studies have been more concerned with pre-task planning than online planning (three in oral production studies and four in written production studies). The findings are as follows. Pre-task planning has more effects on fluency and complexity than on accuracy. Online planning has a little effect on accuracy but no effect on complexity. It has a negative effect on fluency.

2) The majority of planning studies have focused on investigating the effects of task planning on oral production than on written production (26 on oral production and 9 on written production). A comparison between the results from studies on the effects of planning on oral production and those on the effects of planning on written production reveals that planning seems to be more effective in promoting speech performance than writing performance.

3) Very few studies investigated the effects of planning on IL development. Almost all studies reviewed in this chapter examined only the immediate effects of planning on language performance. The only study that did probe into this issue was Bygate's study (2001) on the effects of task repetition, which has been reviewed in section 3.5.2. Bygate's study (2001) was important in that it addressed the issue of the relation between L2 performance and L2 learning. The relation between L2 performance and L2 learning will not be established unless studies are designed to find out if the effects of task planning can be transferred to a new task and if the effects can be retained for a longer period of time (R. Ellis, 2005b). To date, no study has been carried out to investigate the longer-term effect of pre-task planning and online planning. Thus, whether or not pre-task and online planning can facilitate L2 learning remains a question yet to be answered.

4) Though the number of studies that have taken a process-product approach by also looking into what learners actually do during planning is increasing, our knowledge of

learners' planning process is still limited. How their planning process influenced their subsequent writing process was not examined because these studies limited themselves to studying the planning process only. Only when both the planning process and the subsequent writing process are studied can we have the whole picture. Only then can we understand better why and how planning affects language performance. Another limitation of previous studies on the planning process is that learners' planning process has been probed into mainly by using learners' retrospective self-report (e.g. Ortega, 1999; R. Ellis & Yuan, 2004; Nakakubo, 2011). While self-reports provide insight into the mental processes the learners engaged in, they have limitations in that there could be mismatches between what learners reported and what actually took place. Therefore, it is necessary to triangulate self-report data with other types of data, such as learners' planning notes.

### **3.10 Research gaps the present study addresses**

It can be seen from the above summary that there is scanty research on the effects of planning on L2 writing and online planning. The role planning plays in L2 development is still opaque for the lack of empirical study that looks at the transfer of planning effects to a new context. The process of planning has not been adequately investigated and the process of subsequent writing has not been examined. To address these research gaps, the present study looks at the relative effectiveness of pre-task planning and also online planning on the immediate performance in L2 writing and on L2 writing development. Further, it examines the processes learners engage in not only during pre-task planning but also during task completion. The details of how these gaps were attempted to be addressed will be presented in the next chapter.

## CHAPTER 4 METHODOLOGY

### 4.1 Overview

In order to investigate the effects of task planning on L2 writing, a quasi-experimental study was conducted. This chapter addresses the methodology of the study. In the following sections, research questions will be presented first. This will be followed by an overview of the research design. After the overview, the participants, operationalization of planning conditions, research instruments, experimental procedures, data coding and scoring procedures, and data analysis techniques will be described.

### 4.2 Research questions

This study is designed to answer the following questions:

1. Does task planning have effects on immediate L2 writing performance?
  - a. Do learners under the pre-task planning condition outperform those under the no planning condition in terms of accuracy, complexity, fluency, and organization of their written production?
  - b. Do learners under the online planning condition outperform those under the no planning condition in terms of accuracy, complexity, fluency, and organization of their written production?
2. Does repeated practice under planning conditions lead to changes in L2 writing over time?
  - a. Does pre-task planning lead to increased accuracy, complexity, fluency, and organization in learners' written production over time?
  - b. Does online planning lead to increased accuracy, complexity, fluency, and organization in learners' written production over time?
3. What do learners do during planning and writing time?
  - a. What do they focus on during pre-task planning time?
  - b. What strategies do they use to plan their writing?

- c. How much of their plan is applied in subsequent writing?
- d. What are their planning styles?
- e. What processes do they engage in while writing the essays?
- f. How do they perceive their task completion conditions?

Research questions 1 and 2 will be answered by an analysis of quantitative data and Research question 3 will be answered by qualitative analysis of learners' answers to the questions in the post-task questionnaire and the semi-structured interview.

### **4.3 Overview of the experimental design**

The study adopted a between-group quasi-experimental design to investigate the impact of task planning on L2 writing with three task conditions: pre-task planning, online planning, and no planning. Three intact classes served as three groups for the experiment: one group had 10 minutes' planning time before they completed each writing task in 15 minutes (PTP), the second group did not have time to plan before they were given the writing task but had 25 minutes to complete the task (OLP), and the third group completed each writing task in 15 minutes (NP). In total, both PTP and OLP had 25 minutes treatment time while NP had 15 minutes. Four writing tasks (see Appendix D) were used to elicit participants' planned and unplanned output, which were completed over a 4-week period.

At the beginning of the study, participants filled out a questionnaire for demographic information (see Appendix B) and completed a pre-test (see Appendix C for tasks used in testing sessions) to establish a baseline for research. For the tasks used in the testing sessions (i.e. pre-test, posttest, and the delayed posttest), participants were allowed 25 minutes to complete without any pre-task planning time. In the four weeks following the pre-test, they completed one experimental writing task each week. In the week following the fourth experimental task, a post-test was administered. Then a post-task questionnaire (see Appendix E) was administered to collect data on how the planners planned and completed the tasks, and how the non-planners completed the tasks. This was followed by a semi-structured interview (see Appendix F), in which half of the participants were

interviewed. The interview was conducted to obtain more in-depth information on participants' planning and task completion processes and their attitudes towards their task completion conditions. Students' planning notes were used as prompts to stimulate their memories. Nine weeks later, a delayed post-test was administered. This was immediately followed by a post-study questionnaire (see Appendix G) to find out learners' attitudes towards writing in English in general, their feelings about making mistakes, and whether they had realized which linguistic features were targeted. Table 2 shows the overview of the experimental design of the study.

Table 2 Overview of the experimental design

Week 1	Recruit participants + questionnaire for background information
Week 2	Pre-test
Week 3-6	Treatment sessions (4 sessions) PTP: 10 min planning+15 min writing OLP: 25 min writing NP: 15 min writing
Week 7	Post-test + Post-task questionnaire
Week 8	Semi-structured interview
Week 9-17	Break
Week 18	Delayed post-test +post-study questionnaire

## 4.4 Participants

### 4.4.1 Selection of participants and their proficiency level

107 non-English major freshmen from a university in Beijing volunteered to participate in the study. At the time they were recruited, they had just graduated from high school and passed the college entrance examination. According to the National English Curriculum Standard for High Schools (*National English Curriculum Standards*, 2001) in China, a high school graduate should have a vocabulary size of around 2,500 words. The participants of the present study should have met this requirement since the

university is a key university in Beijing and students need to be quite competitive to get in. From the perspective of vocabulary size, this group of learners could be considered to be at the intermediate level of proficiency, in light of McCarthy's (McCarthy, 2007) proposal that 6,000 words is the entry for advanced level.

Upon entering university, all non-English major freshmen of this university sat a placement test, which consisted of four parts: listening comprehension, reading comprehension, grammar and vocabulary, and cloze test. The test, which was designed by Beijing Municipal Commission of Education and Beijing College English Research Association, had been used by many universities in Beijing to stream non-English major students. After the test, the students were put into English classes of different levels according to their scores for the placement test. The participants of the present study were from three English classes of the same level.

Initially, there were 107 students. Two weeks after recruitment, two students dropped out of the university. Two students from NP, ten from OLP, and seven from PTP missed either one test or one or more experimental sessions. Therefore, data from 19 students were incomplete, thus not used in the analysis. Though the three classes were from the same level, there were still differences among them in terms of their mean placement test scores. To minimize the pre-treatment differences between groups as much as possible, one of the two frequently used methods suggested by Dörnyei (2007) was used, i.e. matching participants by equating the groups. This was done by omitting some participants whose scores were either much higher or lower than the rest of the group (i.e. outliers). In the present study, the outliers refer to those whose scores were two standard deviations from the group mean. After excluding the extreme cases, data from 75 participants with 25 in each group were coded and subjected to analysis.

The mean placement test scores for the three groups after trimming were respectively 66 (PTP), 65 (OLP), and 64.5 (NP). A one-way ANOVA test ( $F(2, 72)=1.635, p=.203$ ) revealed that the scores of the three groups were not significantly different from one another, which indicated that their general L2 proficiency levels were about the same. Their pre-test scores, which will be presented in Chapter 6, confirmed this.

#### 4.4.2 Demographic characteristics of participants

All together there were 48 females and 27 males and their ages ranged from 17 to 20. They were all Chinese students who had studied English under a formal instruction condition in schools (i.e. EFL context). None of them had overseas learning experience and they had little opportunity to use English for communicative purposes outside the classroom. Their years of learning English ranged from six to twelve years with an average of 7.8 years. It can be concluded from the above-mentioned information that the participants constituted a fairly homogeneous group in terms of age, educational background, English learning history, English proficiency, and their L1 (see Table 3). The groups differed in the composition of gender, though. OLP and NP had a fairly good balance of male and female students, while PTP had a much bigger female population than male population. This could be a potential confounding variable since gender has been recognized as one of the factors in some studies (e.g. R. Ellis, 1994) to cause variation in language learning. However, how it affects the effects of planning has not been investigated and it is not the focus of the present study either.

Table 3 Demographic information of the participants

Group	Gender		Age	Average Years of learning English	Oversees learning experience
	Male	Female			
PTP	4	21	18.2	8.5	None
OLP	11	14	18.6	7.6	None
NP	12	13	18.4	7.6	None

#### 4.5 Operationalization of planning conditions

##### 4.5.1 Operationalization of pre-task planning

In previous studies, pre-task planning has been operationalized according to four criteria: 1) source of planning (Foster & Skehan, 1999); 2) the amount of time allowed for pre-task planning (Crookes, 1989; Foster & Skehan, 1996); and 3) the inclusion or exclusion of guidance to planners (Foster, 1996; Foster & Skehan, 1996; Sangarun, 2001, 2005).



Source of planning is defined as whether pre-task planning is teacher-led, group-based, or solitary (Foster & Skehan, 1999). In a teacher-led condition, the teacher guides the participants in how to plan their production. In group-based pre-task planning, participants work in small groups and prepare their production collaboratively. Participants in a solitary planning condition work on their own during pre-task planning time. For the present study, it was decided that participants would be asked to do solitary pre-writing planning. This decision was made with a view to avoiding confounding the effects of planning with the effects of instruction or collaboration.

With respect to the second criterion, the amount of time given to planning prior to the task was 10 minutes in most previous studies and was believed to be the optimum length of time (e.g. Crookes, 1989; Foster, 1996). The current study followed other studies and made 10 minutes available for participants to prepare before they started writing the essays in the treatment sessions. During the 10 minutes, they were encouraged to make notes of what they were going to write for the task. As in other studies (e.g. R. Ellis & Yuan, 2004; Foster & Skehan, 1996; Guar-Tavares, 2008; Wang, 2009), when they started writing the note sheets were removed to prevent the planners from copying from their notes so that the pre-task planners had the same amount of actual writing time as the non-planners.

Regarding the third criterion, guidance to planners typically includes two aspects: focus of planning and suggestion on how to make use of the planning time. The focus of planning is typically manipulated through planning instructions and employment of specially designed planning sheets to have learners focus on meaning, form, or on both meaning and form (Sangarun, 2001, 2005). In Sangarun's study (2001, 2005), under the focus on meaning condition, learners were instructed to plan the meaning of their speech and they were guided to generate appropriate information and shape the information of their speech. Under the focus on form condition, learners were guided to plan both vocabulary and grammar for their speech. In the third condition, learners were guided to plan both meaning and form. They were instructed to plan the meaning before they planned the form and were told to integrate their planning of meaning and form.

Sangarun's study (2001) found that focusing on both meaning and form was the most beneficial to the learners' speech performance. Therefore, it was decided that for the present study, learners would be instructed to plan both meaning and form (see Section 4.7.4.1 for instructions for pre-task planners).

As for the guidance on how to spend the planning time, previous studies that investigated the effects of this variable (Foster, 1996; Foster & Skehan, 1996; Skehan & Foster, 1997) influenced the design of the present study. In these studies, pre-writing planning has been operationalized as guided versus unguided planning, or detailed as distinct from undetailed planning (Foster & Skehan, 1996). In the guided or detailed planning condition, participants are given detailed suggestions on how they should make use of the planning time to prepare their writing. In the unguided or undetailed planning, participants can plan at will without any guidance on how to make use of their planning time. These studies did not provide enough evidence to show that guided planning is superior to unguided planning. Therefore, for the present study, participants were given limited guidance on how they should spend their 10-minute planning time (see Section 4.7.4.1).

To sum up, the pre-writing planning condition for the current study was operationalized as follows:

1. The planners were given 10 minutes to plan prior to their completion of the writing tasks.
2. They were instructed briefly to plan both content and language.
3. They were given limited guidance on how they should make use of their planning time.
4. They were left to plan on their own during the pre-task planning time.

#### **4.5.2 Operationalization of online planning**

Online planning has not been extensively studied in previous planning studies. Among the few studies (e.g. D. Li, 2004; Yuan, 2001) that have been found so far to investigate the effects of online planning on written production, Yuan (2001) gave the

online planners unlimited time to compose a story. This on the one hand freed the online planners from time pressure and gave them ample time to do online planning, but on the other hand confounded the effect of online planning with pre-task planning. This is because when learners have unlimited time to compose, they might use the time to do both pre-task planning and online planning. To avoid this problem, D. Li (2004) set a time limit for learners in the online planning group. In her study, the pre-task planning group was given 17 minutes to write the composition and the online planning group was given 30 minutes to compose.

Following D. Li's study (2004), the online planners in the present study were told to begin writing as soon as they finished reading the instructions and they were allowed 25 minutes to complete the essay (the pre-task planners had 15 minutes to write the essay). This was intended to limit the time the online planners had for pre-task planning while giving them more opportunities to do online planning than the pre-task planners.

Efforts were made to differentiate the two planning conditions. It has to be noted, though, that the two planning conditions are not mutually exclusive. The possibility for the learners in the pre-task planning group and the no planning group to engage in online planning could not be eliminated even though they had limited composing time. Vice versa, learners in the online planning group might also use some time to do pre-task planning despite the fact that they were instructed to start writing right after they read the instructions.

## **4.6 Instruments**

### **4.6.1 The writing tasks**

The present study used focused writing tasks (R. Ellis, 2003) to elicit data. This was decided so that the effects of planning on the accuracy of specific language structures as well as the general accuracy could be examined. Ellis (2005b) suggests that to obtain evidence on the effects of planning on language development, a pretest-treatment-posttest design should be adopted and specific linguistic feature(s) should be targeted. An

appropriate target structure would be either a newly learnt structure or a structure that is partially learnt and still causing difficulty for the learners. According to Ellis (2005b), learning takes place if any of the following is evidenced: (1) learners are using linguistic forms they have not used previously; (2) there is an increase in accuracy in the linguistic forms they are already using; (3) the linguistic forms they are already using are used to perform a new function or in a new context; (4) there is an increase in the fluency of using linguistic forms (pp. 27-28). Following Ellis's suggestions, the use of two specific linguistic features---comparative form and the subject-verb agreement were examined in addition to the general accuracy of writing. The following section will provide the reasons for choosing the two target forms and descriptions of them.

#### **4.6.1.1 The target structures**

When selecting target forms, there were three major considerations: 1. The learners should be already using them in language production; 2. The learners have not gained complete control of them so it would be possible to trace linguistic change(s), if any, by looking at if there is any increase in accuracy in their use and if learners can use them correctly in new contexts; and 3. The forms should not be taught during data collection period so the effects of planning could be isolated from the effects of instruction. Comparative forms and subject-verb agreement meet the three criteria outlined above. First, by the time data for the present study was collected the participants had already been using them in language production. This was known through the trial task used when selecting the research site for the main study (see section 4.7.1). Second, the two forms are known to be universally problematic to L2 learners (R. Ellis, 2005a). Why the two structures are difficult for participants of the present study will be detailed in the next two paragraphs. Third, the two structures were not formally taught during the period of data collection. This was verified by participants' answer to Question 5&6 in the post-study questionnaire (see Appendix G). Having presented the reasons for choosing the two target forms, descriptions of the two forms and explanations on why they present difficulty for L2 Chinese learners will now follow.

Subject-verb agreement is a morphosyntactic feature. English grammar stipulates that the subject of a sentence and the verb must agree in number. That is, when the subject is singular the verb has to be singular and when the subject is plural the verb has to be plural. In the present tense, there are two inflectional forms in English (-s and -es). For most lexical verbs when the subject or the entity performing the action is third person singular (*he, she, and it*), an -s or -es has to be added to their base forms. When the subject is plural the base form of the verb is used. For the verb 'be', which is highly irregular, there are three verb forms according to the subject of the sentence in the present tense (i.e. *I am, he is, you/we/they are*) and in the past tense there are two forms (*I/he/she was, we/they/you were*). Though the rule for S-V (subject-verb) agreement is seemingly simple, the mastery of it is notoriously difficult for learners, particularly for those whose native language does not use grammatical morphology to mark number, as in the case of Chinese. This lack of grammatical morphology leads to differences in language processing between native speakers of English and Chinese speakers (Chen, Shu, Liu, Zhao, & Li, 2007; P. Li, Bates, & MacWhinney, 1993; P. Li, Jin, & Tan, 2004). In a recent study on morphosyntactic processing of S-V agreement by Chinese L2 learners of English, Chen et al. (2007) found distinct differences between proficient Chinese L2 learners and native English speakers in the processing of S-V agreement. This led them to conclude that S-V agreement presented significant difficulty to Chinese learners of English even when they had become proficient in L2. Another factor that adds to the learning difficulty of S-V agreement is that the inflectional -s added at the end of the verb in the case of third person singular subject is semantically redundant and phonologically nonsalient. For the reasons outlined above, S-V agreement was chosen to be one of the target structures for the present study. S-V agreement for all verbs in the present and the verb 'be' in the past tense was examined.

The other target structure selected to be under investigation was the comparative form. Comparative expressions with adjectives, adverbs, and nouns are examined. In describing the features of comparative forms, Quirk et al. (1972) pointed out that "The essential feature of a comparative construction, in broad grammatical terms, is that two

propositions, one expressed by the main clause and one by the comparative clause, are compared with respect to something they have in common” (p. 765). A typical comparative construction is: noun X is (comparative adjective) than Y. However, it should be noted that in comparatives used by native speakers 70 percent are constructed without a *than* clause (George, 1972). The form of the comparative adjectives and adverbs is decided by prosody. In most cases, monosyllabic adjectives and adverbs can form their comparison by inflection (-er). Disyllabic adjectives that end with -y, -ow, -er, or -le can also take inflectional forms. Adjectives and adverbs that fall outside these four categories (with a few exceptions, e.g. common, handsome, polite) and multisyllabic adjectives and adverbs usually have to take the periphrastic form (*more*). Comparative forms present difficulty for learners in that the formation of comparative forms requires inflectional or phrasal modification of the adjective or adverb and involves the construction of the whole clause. The fact that the inflectional and periphrastic markers of comparative forms are sometimes semantically redundant makes the form more difficult to learn.

In short, S-V agreement and comparative forms were chosen to be the target structures for the present study for the reasons that learners should be able to use them in their writing but might commit errors in them and that learners would not receive formal instruction on them during the data collection period.

#### **4.6.1.2 Task design**

The design of tasks for the present study followed two principles: 1) the tasks should be at an optimum level of difficulty because very simple tasks may not reveal the potential value of cognitive strategies (Kellogg, 1987) and very difficult tasks might be hard for the participants to complete; 2) the tasks should be able to elicit a sufficient number of occurrences of the target structures and a sufficient amount of output. Based on these principles, it was decided that learners would be asked to write compare/contrast persuasive letters. This was because the learners had practiced letter writing in high school in preparation for the college entrance exam though they had not been taught

formally how to write essays of this genre. The difficulty caused by participants' unfamiliarity with the particular genre and the cognitive demands of persuasive essays was alleviated by the topics of the tasks. The topics used were ones they would be familiar with so they could draw from their own experiences when writing. Moreover, the topics were of a general nature, instead of being too academic. This could give equal opportunities to learners of various academic backgrounds to produce good essays. To further lessen the cognitive demands, contextual support was given for each task (see Appendix C and D). The contextual support included information on the objects of comparison, provided either in the form of a table or a bulletin list, topic related pictures for one of the tasks, and a list of useful words or expressions (most of them are adjectives) provided for each task. The provision of information on the objects under comparison could ease the difficulty of idea generation on the part of the learners. The list of useful words and expressions provided for each task might lend linguistic support to the learners, suggest possible ideas for inclusion, and induce greater use of comparative forms.

#### **4.6.1.3 Description of the tasks**

Seven writing tasks with features described above were designed to elicit the target structures. Four of them were used in the treatment sessions and three were used in the testing sessions. They were all letter writing tasks, which required learners to persuade the reader(s) to agree on the decisions they had made.

For example, Task 2 for the treatment sessions asked the participants to imagine themselves to be a human resource manager of a travel agency, who was selecting a new employee for the agency. Information on two candidates was provided on the task sheet. The participants were required to compare the two candidates and decide which one to hire. They needed to write a letter to the other members of a selecting panel to convince them that the person they had chosen was the right one for the job. The other three tasks used in the treatment sessions had similar designs (see Appendix D).

For each task, the participants were instructed to use as many different adjectives as possible. This was to encourage greater use of comparative forms with different

adjectives. To control for differences in performance caused by different tasks, a split block design was used. For all four treatment sessions, Tasks 1, 2, 3, and 4 were administered during each treatment session with one quarter of each group completing one of the four tasks.

Tasks used for the testing sessions were of the same type and similar difficulty level with those used in the experimental sessions but with different topics (see Appendix C). As in the treatment sessions, the split block design was adopted in test administration to control for possible differences in performance caused by different tasks. For all the testing sessions, all three tasks were used during each session.

#### **4.6.1.4 Task complexity**

Task complexity will be analyzed with the two complexity dimensions described in Robinson's Triadic Componential Framework (Robinson, 2001, 2005, 2007, 2011) since it provides a model for a detailed analysis of tasks in a principled way. As has been discussed in Chapter 2, Robinson distinguishes two dimensions when discussing task complexity: *resource-directing* and *resource-dispersing* (ibid).

*Resource-directing* elements refer to task variables that place conceptual or cognitive demands on learners while *resource-dispersing* elements refer to variables that create favorable task conditions for the learners. For the sake of clarity, categories of task variables along the two dimensions mentioned above are going to be listed and the writing tasks of the present study will be analyzed in relation to the categories. Task variables in the *resource-directing* subcategory include: 1) whether the task requires reference to events happening now (Here-and-Now) or in the past (There-and Then); 2) whether the task requires reference to few or many elements (e.g. making a choice out of two options is simpler than making a choice out of 5 options); 3) reference to a spatial location, where easily identifiable landmarks can be used or not; 4) whether the task requires simple transmission of information vs. figuring out causal relationships between events; 5) simple transmission of information vs. reasoning about other people's intentions and beliefs, etc. and 6) whether the task requires the speaker/listener to take



just one first-person perspective (i.e. *I*) or many second and third person perspectives (e.g. *you* and *s/he*) (Robinson, 2007). Increase in demands along this dimension would place greater conceptual demands on learners.

Judged by the criteria outlined above, the writing tasks used in the present study place relatively low conceptual demands on two dimensions and high demands along two of the six dimensions listed above on learners. First, they are the Here-and-Now type of tasks. According to Robinson's framework, the Here-and-Now type of tasks is less cognitively demanding than the There-and-Then type of tasks, since it requires the use of present tense rather than past tense that is needed for doing the There-and-Then tasks. Second, the writing tasks of the present study require participants to refer to only two objects that they were comparing when persuading the intended reader to agree with them. They are less demanding than tasks which require reference to many elements. Third, tasks of the present study require not only transmission of information but also causal reasoning because participants need to provide reasons for their standpoint to persuade the intended reader. In this aspect, the tasks are demanding. With regard to the perspectives learners had to take in writing, first-person, second-person, and third-person perspectives sometimes had to be used, thus making the tasks demanding (ibid). For example, in Task 1 of the treatment sessions participants had to use all three perspectives by referring to the intended reader as *you* (e.g. *you were unhappy about the baby girl*), the writer himself/herself as *I* (e.g. *I think...*), and the baby girl as *she* (e.g. *she is a gift from God*).

In contrast to the resource-directing variables are those *resource-dispersing* variables that make increasing performance demands on learners' attentional and memory resources. Task variables in this dimension include 1) whether planning time is available or not; 2) whether background information needed for task completion is provided or not; 3) whether it is a single-demand task or a dual/multiple-demand task (e.g. doing a grammaticality judgment task is a single-demand task and doing a grammaticality judgment task and a tone discrimination task simultaneously is a dual-demand task); 4) whether there is a clear task structure or not; 5) whether the task requires one or few steps

vs. many steps to complete it, and 6) whether it is necessary to follow a strict sequence of steps.

According to the resource-dispersing framework, the tasks used in the present study did not put a high performance demand on the part of learners for the following reasons. First of all, planning time was provided for the two experimental groups. This made the tasks less demanding for the planning groups. Second, background information was provided on the task sheets for all participants to compare two objects. Third, the tasks only required participants to write letters to persuade the reader(s) to agree with their decisions, therefore they can be regarded as single-demand tasks. Regarding the fourth category, a task structure was provided for some learners, although this was not intended. As has been explained, information on the objects under comparison was provided on the task sheets. Though the information was arranged in random order, participants could choose to follow the order to compose their essays. In this sense, a task structure was provided though it was not desired by the researcher.

In summary, the tasks used in the present study were cognitively demanding along two of the resource-directing dimensions but the cognitive demands were reduced by some resource-dispersing elements. The table below summarizes the features of the tasks in the present study.

Table 4 Features of the tasks in the present study

<b>Robinson’s framework</b>	<b>Tasks for the present study</b>
Resource-directing variables:	
+/- Here-and-Now (reference to events happening now or in the past)	Here-and-Now ↓
+/- few elements (Reference to few or many elements)	Few elements ↓
+/- spatial reasoning (if an easily identified landmark can be used as reference)	Not applicable
+/- causal reasoning (simple transmission of information vs. causal reasoning required)	Need causal reasoning ↑
+/- intentional reasoning (simple transmission of information vs. reasoning about other people’s intentions and beliefs)	Not applicable
+/- perspective-taking (if just first-person perspective, or many second and third-person perspectives is required)	Require use of first-person, second-person, and third-person perspective ↑
Resource-dispersing	
+/- planning time (if planning time is provided)	Planning time provided (for planners) ↓
+/- background information (if background information is provided)	Background information provided ↓
+/- single task (if it is a single demand or multiple demand task)	Single demand task ↓
+/- task structure (if a task structure is provided)	Task structure provided ↓

Note: ↓=cognitive demand is reduced; ↑=cognitive demand is increased.

#### 4.6.2 Questionnaires

To gather information on the demographic characteristics of participants, their responses to the opportunities to plan, and their general attitudes towards writing in English, questionnaires were used in the present study because they are “versatile and [is] capable of gathering a large amount of information quickly in a form that is readily processible” (Dörnyei, 2007, pp. 101-102). In the present study three paper and pen questionnaires were used. The first one which was designed for collecting background information on the participants contained factual questions to ascertain the demographic characteristics of the participants (see Appendix B). The second was developed for

providing information on how learners went about doing the tasks, what the planners did during planning time, and how learners thought of the tasks and task completion conditions. It used both open-ended and closed-ended questions, with open-ended questions aiming at getting responses from participants regarding their processes of planning and essay writing. Closed-ended questions were mostly attitudinal and factual questions asking what they thought of the tasks, how they felt when doing the tasks, and how much of the planned content and language was actually used in writing, etc. (see Appendix E). The third questionnaire contained mostly open-ended questions asking learners to report on their attitudes toward writing in English in general, how much importance they attached to accuracy when writing, and whether they realized that comparatives and S-V agreement were targeted (see Appendix G). Since data elicited from the third questionnaire does not relate directly to any of the research questions, it will not be reported and discussed in detail.

The first questionnaire was completed a week before the pre-test, right after the participants signed the consent forms. It took them five minutes to complete. The second one was completed immediately after the post-test. There was no time limit set for learners but they all finished the questionnaire in about 15 minutes. The third questionnaire was filled out right after the delayed post-test. Again, there was no time limit but the participants all finished in around eight minutes.

#### **4.6.3 Semi-structured Interview<sup>8</sup>**

While having the advantage of being capable of collecting large amounts of information quickly, the questionnaire is “unsuitable for probing deeply into an issue and it usually results in rather superficial data” (Dörnyei, 2007, p. 115). To probe into the cognitive processes learners engaged in when planning, concurrent online and retrospective off-line protocols (e.g. retrospective interview) were considered better tools since they are “the closest reflection of the cognitive processes” (Ericsson & Simon, 1993,

---

<sup>8</sup>These types of interview are classified as retrospective off-line protocols (Read, 2000) as they were carried out after a complete process of treatment.

p. 16) and they don't modify the cognitive processes being reported. However, off-line protocols cannot provide as much of thought processes as possible and (Taylor & Dionne, 2000) data might be lost because of recall failure. One advantage off-line protocols have over concurrent reports is that they are less reactive because the thought processes are not affected by the procedure in any way (Ericsson, 2002). Because of this, it was decided that semi-structured retrospective interviews would be conducted to complement information that could not be gathered by the post-task questionnaire.

In the interview, the participants were asked to answer questions on their process of planning, the application of their plans, their writing process, and their attitudes towards planning, etc. (see Appendix F). With respect to the time retrospective interviews are best conducted, Dörnyei (2007) suggested that the interval between the task and the retrospective interview be kept as short as possible, so memories of what has been done during the process of task completion are still fresh in the participants' minds, and can thus be recollected easily. In the light of Dörnyei's suggestion, the semi-structured interview was conducted in the week after the post-test. This was not ideal but it was the earliest possible time for the participants. During the interview, learners' planning notes and answers to the post-task questionnaire were used as prompts for recall. The procedure of the interview will be described in more detail in the next section.

## **4.7 Procedures for data collection**

### **4.7.1 Selection of research site and participants**

Data was collected from October 7<sup>th</sup>, 2009 to March 5<sup>th</sup>, 2010. Before data collection started, a trial task, which was similar to the experimental tasks used in the study, was given to 15 students from three universities (five from each) to find out which university had the right population of students for the present study in terms of their level of English writing proficiency. Students, who were able to complete the task, produced a sufficient amount of output, and who created a sufficient number of obligatory occasions for use of the target structures but demonstrated incomplete acquisition of these, were regarded as

suitable for the present study. The results of the trial showed that students from one of the universities were most appropriate for the study. Then the HOD of the university was contacted and approval for the study was obtained. After recruitment of participants and a check on availability of necessary facilities were completed, the data collection began. The five students who completed the trial task were not in the three classes who participated in the present study.

In sum, the university, whose students were identified in the trial as appropriate for the present study, was selected to be the research site for two reasons: 1) it had EFL learners of the English proficiency level appropriate for the study (as ascertained by the trial) and 2) the head of its Foreign Language College agreed to support the study.

The procedures for recruiting the participants were as follows. The head of the Foreign Languages College was contacted first and she then introduced the researcher to the class teacher who was to assist with recruitment of participants. The researcher was led to each classroom at the beginning of the class. The students were informed of the purpose of the study, what they would be required to do for the study, the time requirements, and the benefits they would get from participation. The researcher left the information sheets in the classroom for the students to read during the break and asked those who were interested in participating to sign the consent forms. After class, the participants signed the consent forms and submitted them to the researcher.

#### **4.7.2 Assigning participants into research conditions**

The participants were all freshmen from four departments (law, finance, accounting, business management) of the university, who were put into three classes for English lessons. They had two hours English reading classes, two hours listening and speaking classes each week, but no writing classes. The three classes were arbitrarily assigned to the pre-task planning group (PTP), the online planning group 2 (OLP), and the no planning group (NP) according to the time they had English classes. The 8 am-10 am class was assigned to PTP; the 10am-12am class was OLP; and the 2pm-4pm class was NP. The three groups met the researcher once a week separately to do the pre-test, the

experimental tasks, the post-test, and the delayed post-test.

#### **4.7.3 Procedures for the testing sessions**

Three writing production tests (see Appendix C) were designed for the study. The pre-test was administered one week prior to the treatment sessions. The post-test was administered one week after the treatment sessions and the delayed post-test was conducted nine weeks after the post-task semi-structured interview (see section 4.3 for the timetable). All tests were administered by the researcher.

The tests took place during timetabled class hours so the administration was under time constraint. Each testing session was not allowed to exceed 30 minutes. For this reason the participants were required to complete each task in 25 minutes. How to make use of the time and whether to use a dictionary or not were up to the learners to decide for this was the common practice in writing tests in Chinese universities.

Prior to the pre-test, a sample essay was written by the researcher to exemplify the type of essay that was expected of them. This was done for the fact that the participants had not been taught how to write persuasive compare/contrast essays and might not know how to compose such essays. The sample essay compared two restaurants for the purpose of persuading a friend to choose one of them to go to for dinner. It demonstrated to the participants what a persuasive compare/contrast essay was like. Each participant was given one sheet, on which the sample essay was written. They had five minutes to read it and ask questions, if there was any. After five minutes, the sheet was removed to prevent participants from copying from it.

At the beginning of each testing session, the researcher handed out the task sheets and told the participants that they would have 25 minutes to complete the task. Then she waited in the classroom till the end of the testing session. When 25 minutes were up, the participants were asked to stop writing and hand in the essays.

#### **4.7.4 Experimental procedures**

Three experimental procedures were used in the study: one for PTP, one for OLP, and

one for NP.

#### 4.7.4.1 Procedures for the pre-task planning group

The participants in the pre-task planning group worked with the researcher according to the following steps.

1. The researcher explained to them in Chinese what they would be asked to do as follows:

*You will be given a writing task to do. Before you start writing, you will have 10 minutes to prepare for the task. You will be given a sheet of paper to jot down your notes. After 10 minutes, you will have 15 minutes to write.*

2. The researcher then gave the instructions written below and passed out the note sheets and task sheets.

*This is the writing task you are required to do. After reading the instructions, you will have 10 minutes to prepare for your essay. Here is one sheet of paper on which you can make notes. When you are preparing, please try to think of the vocabulary and grammar you may use in your essay as well as how to make your essay more convincing. Once you begin to write your essay, I will take away your note sheets. Now, please read the instructions for the task and start your preparation. I will notify you when the preparation time is over.*

3. When the 10-minute planning time was over, the researcher asked the participants to stop planning, hand in their planning notes, and start to write the essay. The instruction the researcher gave was as follows:

*Now it is time for you to write the essay. You will have 15 minutes to complete it. Please use a pen and if you need to make corrections just cross out the words you don't want.*

After giving instructions, the researcher went to the back of the room and waited for the participants to finish. When the 15-minute writing time was over, most of the participants naturally stopped writing. If any of them indicated that they needed a little extra time (no more than 1 minute) to finish, that was allowed but the actual time spent



on the essay was noted down on the sheet.

#### **4.7.4.2 Procedures for the online planning group**

The online planning group followed the steps below:

1. The researcher explained to them in Chinese what they would be asked to do as follows:

*You will be given a writing task to do. For this task, you will have 25 minutes to write. Please write with a pen and cross out the unwanted words when you need to make corrections.*

2. When the 25-minute writing time was over, the researcher asked the participants to stop writing and hand in the essays. Sometimes some students finished writing before the researcher asked them to stop. In this case, they were allowed to hand in their essays early but the actual time spent on the essay was noted down on the sheet.

#### **4.7.4.3 Procedures for the no planning group**

The no planning group followed the steps below:

1. The researcher explained in Chinese what they would be required to do as follows:

*You will be given a writing task to do. For this writing task, you will have 15 minutes to write. After 15 minutes I will collect your essays. Please write with a pen and when you want to make a correction just cross out the unwanted words.*

2. When 15 minutes were over, the researcher asked them to submit the essays. If anyone indicated that s/he needed extra time to finish, no more than two minutes were given and the actual time spent on the essay was noted on the sheet.

#### **4.7.5 Procedures for the post-task questionnaire and the post-study questionnaire**

Immediately after the post-test, the post-task questionnaire was handed out to all participants. During the time the participants were filling out the questionnaire, the researcher was in the room to answer questions from the participants. The post-study

questionnaire was administered right after the delayed post-test. Both questionnaires were written and completed in Chinese.

#### **4.7.6 Procedures for the post-task interview**

The interviews were conducted in the week following the post-test. On the same day, the participants had their regular final oral exam in the university. For the oral exam, the students had to go into the exam room one by one to take the test. An exam assistant was there to help with the administration. The order in which the students took the exam had been arranged by the class representatives in advance and the name lists were obtained from them. As has been described in section 4.3, half of the participants were interviewed. The researcher picked students with odd numbers on the list to be the interviewees. The room across the corridor from the oral exam room was allocated to the researcher to be the interview site. The exam assistant told the interviewees that they should come to the researcher to be interviewed after their oral exam. Each interview lasted for about 10 minutes and was conducted in Chinese. All interviews were audio recorded.

### **4.8 Data coding and scoring procedures**

#### **4.8.1 The writing tasks**

Multidimensional features of learners' written production were analyzed. To be more specific, learners' essays were analyzed in terms of accuracy of the target forms, overall language accuracy, syntactic complexity, lexical complexity, fluency, and organization of the essays. To carry out the analysis, the essays were typed into the computer and coded for T-units, clauses, errors, error-free clauses, and occurrences of the target structures. Two raters were recruited to grade coherence and cohesion of the essays using the scoring scale of the IELTS. An online software program (Coh-Metrix) was also used to provide automated scores for coherence and cohesion of the essays. The following sections provide a more detailed description of these measures.

#### 4.8.1.1 The accuracy measures

Accuracy refers to “how well the target language is produced in relation to the rule system of the target language” (Skehan, 1996b, p. 23). One way of measuring accuracy is by looking at the specific linguistic structures used in obligatory occasions. This measure, however, is not able to represent the overall ability of the learners to use the second language. To overcome this problem, general measures such as percentage of error-free clauses or the number of errors per 100 words can be used. These measures are able to provide a more general measure for learners’ overall accuracy in using an L2. Skehan and Foster (1999) contended that “a generalized measure of accuracy is more sensitive to detect differences between experimental conditions” (p.229). For this reason, percentage of error-free clauses and the number of errors per 100 words have been widely used (Ellis & Barkhuizen, 2005). To better measure accuracy, it is suggested that both specific measures and general measures are used (Ortega, 1999; Ellis & Barkhuizen, 2005). Following this suggestion, the current study made use of both general measures for accuracy and a measure for accuracy of the specific target linguistic structures.

##### *Overall accuracy*

Two measures for overall accuracy were used in the present study: *percentage of error-free clauses* and *number of errors per 100 words*. The *percentage of error-free clauses* was calculated by dividing the number of clauses without any errors by the total number of clauses multiplied by 100. When marking errors in students’ essays, errors in syntax, morphology, word order, and words which were “nonexistent in English, or indisputably inappropriate” (Skehan & Foster, 1997, p. 195) were all counted, while errors in spelling and punctuation were ignored. Percentage of error-free clauses could not give a complete picture of language accuracy because some clauses may contain more than one error. This problem can be solved by measuring *errors per 100 words*, when every error is taken account of. In the present study, this measure was calculated by dividing the total number of errors by the total number of words multiplied by 100

(Foster & Skehan, 1996; Mehnert, 1998).

### ***Measures for accuracy of the target structures***

The current study selected comparatives as one of the target structures to be examined in students' writing. An obligatory occasion analysis (Ellis & Barkhuizen, 2005) was used to measure learners' use of comparative forms. The formula used to calculate the percentage of accurate use of comparative forms is as follows:

$$\frac{\text{n correct suppliance in contexts}}{\text{Total obligatory contexts}} \times 100$$

The other target form examined in the study was S-V agreement. For all the writing tasks, the participants had to use present tense in most of the cases, yielding plenty of chances for the researcher to look at S-V agreement in their essays. For each essay, lexical verbs with present tense, auxiliary 'have' in the present perfect aspect, and both present and past tense of the verb BE used as either an auxiliary or main verb were marked and judged to see if they agreed with their subjects. Then an obligatory occasion analysis was carried out, using the formula above, to obtain the percentage of the correctly used verbs. For both target forms, only grammatical accuracy was examined. Errors in spelling were ignored.

#### **4.8.1.2 The complexity measures**

According to Skehan (1996b), complexity "concerns the elaboration or ambition of the language which is produced" (p.23). In previous planning studies, it has been measured mainly by syntactic complexity while lexical complexity has been largely ignored (Skehan, 2009b), as was evident in the review of planning studies in Chapter 3. The present study decided to measure both the syntactic complexity and lexical complexity of learners' written texts.

#### ***Syntactic complexity***

Syntactic complexity refers to the "range of forms that surface in language production and the degree of sophistication of such forms" (Ortega, 2003, p. 492). Foster

and Skehan (1996) contend that complexity can reflect the extent to which learners are using their ‘cutting edge’ interlanguage (Skehan, 1996a) and is associated with their willingness to take risks in using language that they have not gained full mastery. It is an important index of language development. In L2 writing, specifically, it has been used to examine differences in the written production of L2 learners across proficiency levels and to investigate task-related variation in L2 writing (Ortega, 2003). Syntactic complexity has been measured by looking at the amount of subordination, the length of production units, and variety and/or sophistication of linguistic forms (Norris & Ortega, 2009). They measure different dimensions of complexity and each has numerous metrics. Commonly used measures for the amount of subordination are mean number of clauses per T-unit/C-unit/AS-unit and mean number of subordinate clauses per total clauses. For length-based metrics, it is important to choose the discourse unit most appropriate for the data to be analyzed. Since the present study used written data, T-unit would be the most appropriate. Variety dimension could be measured by use of some specific linguistic features (e.g. different verb forms in Yuan & Ellis, 2003), mean number of verb arguments, and type-token ratio, etc. (R. Ellis & Barkhuizen, 2005). In a study that looked at the strength of the relation between global measures of L2 syntactic complexity and language development and language proficiency, Wolfe-Quintero et al. (Wolfe-Quintero, Inagaki, & Kim, 1998) concluded that mean length of T-unit, mean length of clause, number of clauses per T-unit, and dependent clauses per total clauses were the most satisfactory measures. Therefore, the current study used *mean length of T-unit* (MLT) and *number of clauses per T-unit* (C/T) to measure global syntactic complexity of the learners’ essays.

To calculate *mean length of T-unit*, T-units should be identified first. A T-unit is defined as “one main clause plus whatever subordinate clauses happen to be attached or embedded with it” (Hunt, 1966, p. 735). A simple T-unit consists of an independent clause including a subject, a finite verb, and any other elements such as adverbial, direct object or complement (Quirk & Greenbaum, 1973). In contrast, a syntactically complex T-unit consists of a main clause and its dependent clauses. Coordinate clauses are not

considered dependent clauses, thus a sentence with a main clause and a coordinate clause was counted as two T-units. The following examples of simple and complex T-units were taken from the data of the current study.

(1) Obviously, Job 2 offers more salary than Job 1.

(2) I would like to choose Jacqueline to be our staff, because she is more experienced than Dave.

(3) These mean the university in UK is more famous than the Chinese university \and it has better qualified education.

Example (1) is a simple T-unit which has a main clause. Example (2) is a complex T-unit consisting of a main clause and a subordinate clause. Example (3) has two T-units separated by \ .

The *mean length of T-unit* is calculated by dividing the total number of words used in the essay with the number of T-units. The bigger the resulting number is, the longer the average length of T-unit and the greater the syntactic complexity of the essay.

To calculate the *number of clauses per T-unit*, clauses need to be identified. A clause is a pair or group of words that consists of a subject and a predicate. It could be a finite clause or non-finite clause depending on whether it contains a finite or non-finite verb. A clause can also be subordinate or coordinate. The following are some examples to show the relation between a clause and a T-unit. The mark for the boundary of a T-unit is a slash \ and that for a clause is a double colon ::.

(4) We'll discuss more details:: when I'll be at home.\

(5) I strongly believe:: that the second job will be more difficult than the first\ ::but it has many features as well.\

Example (4) is a complex T-unit with one main clause and a subordinate clause (a subordinating adverbial clause). Example (5) has two T-units and three clauses, one main clause plus one subordinate clause and one coordinate clause.

By dividing the total number of clauses with the total number of T-units, we can get the *number of clauses per T-unit*. The higher the number is the greater the syntactic complexity of the essay.

### ***Lexical complexity***

A variety of measures for lexical complexity have been used in previous studies on planning: lexical variety/richness, lexical density, and lexical sophistication (Read, 2000). The present study adopted two measures for lexical complexity, i.e. *lexical variety* and *lexical sophistication*, since they reflect different aspects of vocabulary use as will be explained in the following paragraphs.

*Lexical variety* refers to how varied a speech or written text is in terms of lexical items. It has been measured by type-token ratio (TTR). “Type” refers to the number of different words used in language production while “token” is the total number of words produced. The value of type-token ratio is normally between 0 and 1. The higher the ratio the more diverse and richer the lexis is. However, there is a problem with this measure. The ratio of word-types to word-tokens falls as the text gets longer because the longer a text is the more likely some words (e.g. *a, the, is*) are repeated (Meara & Miralpeix, 2004b). Therefore, shorter texts are easier to obtain a high type-token ratio than longer ones (R. Ellis & Barkhuizen, 2005). An alternative measure, which could compensate for this problem, is *mean segmental type-token ratio*. To calculate this, a text is first divided into segments (for example, 40 or 50 words each) and the type-token ratio of each segment is computed. The mean score for all the segments is then calculated (R. Ellis & Barkhuizen, 2005).

The *mean segmental type-token ratio* in this study was calculated with a computer program: D\_Tools (Meara & Miralpeix, 2004a). This program was developed based on the work of Malvern and Richards (e.g. Malvern, Richards, Chipere, & Durán, 2004). Malvern and Richards tackled the problem caused by simple type-token ratio measure outlined above by generating a set of TTRs for the text. These TTR values can be presented as a curve, which can then be summarized by a single parameter, D. The process of computing D value is complex. First, 100 samples of 35 words randomly selected from a text are generated. For each sample, the type-token ratio is calculated, and then the mean TTR for all the samples is computed. Next, 100 samples of 36 words

are generated. With the same process, the mean type-token ratio of all the samples is computed. This process is repeated with 100 samples of 37 words, 38 words, and so on up to 100 samples of 50 words. The result is a set of 16 mean segmental TTR values. “The D value for the text is calculated by matching these values to a series of curves generated by the formula:

$$\text{TTR} = D/N * ((1 + 2*N/D)^{1/2} - 1)$$

The value of D which produces the best fitting curve is taken to be the value which best describes the lexical richness of the text” (Meara & Miralpeix, 2004b, p. 2). A lower value of D indicates that the text contains a lot of repetition and is not as lexically rich as ones that have a higher D value.

Another measure of lexical complexity is *lexical sophistication*. It refers to the percentage of ‘advanced’ or sophisticated words (i.e. lower frequency words) in the text (Laufer & Nation, 1995). It is calculated by dividing the number of ‘advanced’ words by the total number of tokens in the text. In the present study, lexical sophistication was calculated by P\_Lex (Meara, 2001), a computer program specifically designed to assess the lexical complexity of texts used in language teaching. The program takes word frequency as the base for classification and ‘advanced’ words in this context means words that are not found in the high frequency words listed in P\_Lex dictionary file, which is based on Nation’s word lists (Nation, 1988). Words that are among the most frequent words accounting for 50 % of all the words occurring in running text in English are coded ‘0’. Words that are on the list of the next 1000 most frequently used words are coded ‘1’ while words that are not on Level 0 and Level 1 lists are coded ‘2’. Numerals and proper names are coded as ‘1’. One of the advantages of P\_Lex is that it not only works with long texts but produces workable figures for relatively short texts as well, such as texts produced by language learners (Meara & Bell, 2001). When analyzing a text, P\_Lex chops it into segments of 10 words each and counts the number of ‘difficult’ words in each segment. Then a profile is constructed that shows the proportion of segments with 0 difficult words, segments with 1 difficult words, and so on up to 10. P\_Lex calculates the theoretical Poisson curve which most likely matches the actual data produced from the



text and a lambda ( $\lambda$ ) value is generated to indicate the text lexical sophistication. Lambda values range from about .5 to 4.5 and the bigger the figure is the more lexically sophisticated a text is.

#### **4.8.1.3 Measures for fluency**

In the literature on the effects of task planning, fluency is defined as rate of production (Skehan, 1996a) and the capacity to produce language in communication without undue pauses or hesitation. This measure is more commonly used in spoken language. However, it can also provide information on how fluent a writer is and has been used in writing research (e.g. Kellogg, 1988; Kellogg, 1990). In previous studies, fluency has been measured by rate of production (temporal fluency) and dysfluency (repair-fluency), which is a term used in contrast with fluency to refer to interruptions in the flow of production. In the present study both temporal fluency and repair-fluency were measured.

Temporal fluency was measured by the *number of words per minute* and was calculated by the total number of words against the total amount of time used for writing. Repair-fluency (dysfluency) was measured by the ratio of number of crossed out and added words to the total words. It was calculated by dividing the number of crossed out and inserted words by the total number of words written in a text.

#### **4.8.1.4 Measures for the organization of the essays**

In addition to the lexical and syntactical level linguistic features, the organizational features of the essays were also analyzed in this study. Organization refers to how different parts of a text are arranged and structured. It can be discussed at two levels: macro level and micro level. The macro level organization of essays concerns the structure, i.e. the essays should have an introduction, a body, and a conclusion properly sequenced. To determine the measures to be used in the analysis of organization, a preliminary analysis was conducted. In the preliminary analysis macro level organization

was examined in terms of the presence of the introductory and concluding paragraphs and the results showed that essay structure was not a feature that could differentiate better essays from less well written essays produced by the participants of the present study. Thus, it was not pursued in later analysis. At micro level, coherence and cohesion of texts are of primary importance to readers, because they link the ideas and create a meaningful text (Lee, 2002). A text with poor cohesion and coherence seriously impedes readers' understanding of the text. The importance of cohesion and coherence necessitates an evaluation of them in the analysis of text organization.

Before discussing the means to evaluate them, it is necessary to look at the definitions of them. Coherence is an abstract and fuzzy concept (Connor, 1990; Roberts & Kreuz, 1993) but generally it is understood as "the relations that link the ideas in a text to create meaning" (Lee, 2002) and it "pertains to discourse-level features or macro-propositions that are responsible for giving prose its overall organization" (Chiang, 1999). Some scholars contend that coherence not only resides in the text but rather is a result of interaction between the reader and the text (e.g. Yule, 1996, cited in Watson Todd, Thienpermpool, & Keyuravong, 2004). Overall, it can be said that coherence is related to sentence connections and paragraph structure and concerns how the reader interprets the text.

Cohesion is "explicit linguistic devices used to convey information, specially the discrete lexical cues used to signal relations between parts of discourse" (Reid, 1992, p. 81). It is seen as an important contributor to coherence and even considered as a sub-element of coherence (ibid). In some L2 writing studies, coherence has been described in terms of cohesion and the ordering of information in the text. There is still no consensus regarding the definition of coherence and how it is best assessed. Since the present study does not aim to contribute to the discussion about coherence and cohesion and they are so closely connected, they were not measured separately. Instead, they were assessed as a whole.

To get a general picture of how well the essays were organized, two ESL teachers from a university, one of whom had 10 years of experience teaching writing and the other

12 years of experience of teaching writing, were asked to rate the coherence and cohesion of the essays using the public version of IELTS (International English Language Testing System) writing band descriptors as the marking criteria.

In addition to the subjective rating of coherence and cohesion of the essays, an online computer program for analyzing texts was used---Coh-Metrix 2.1 (McNamara, Louwerson, Cai, & Graesser, 2005). This is a recently developed sophisticated computer tool that draws on recent advances in psycholinguistics, discourse processing, corpus linguistics, and computational linguistics (Graesser, McNamara, Louwerson, & Cai, 2004; McNamara, Louwerson, & Graesser, 2002) to analyze texts on various dimensions. It provides automated scores to indicate text cohesion and coherence at both local and global levels (Louwerson, 2002; McNamara, Kintsch, Songer, & Kintsch, 1996) as well as indices for syntactic complexity, readability, and other language characteristics. The tool has been used in a number of studies to measure coherence and cohesion of written texts (Kormos, 2011; Liang, 2006) and speeches (Hultgren, 2009).

Indices from Coh-Metrix (see Appendix H) with theoretical connections to text cohesion and coherence were selected on the basis of the literature (Halliday & Hasan, 1976; e.g. Kintsch & van Dijk, 1978). They included incidence of connectives, lexical overlap, semantic coreferentiality, and anaphoric reference, which are discussed in more detail in the following paragraphs.

*Connectives.* Connectives are measured in Coh-Metrix on two dimensions. The first dimension contrasts positive (e.g. *also*, *moreover*) versus negative (*however*, *but*) connectives and the second dimension relates to the particular conceptual categories of cohesion (e.g. temporal, locational, causal, referential, and additive) proposed by Halliday and Hasan (1976) and Louwerson (2002). For example, *moreover* and *also* are positive additive connectives; *however* and *but* are negative additive; *after* and *before* are positive temporal; *until* is negative temporal, and *because* and *so* are causal connectives. Connectives are important devices for linking ideas. In the present study incidence of all connectives, incidence of temporal, causal, additive connectives, and logical operators (e.g. and, or, if) were used.

*Lexical overlap.* Coh-Metrix assesses four types of lexical overlap: noun overlap, argument overlap, stem overlap, and content word overlap. “For any two sentences  $s_1$  and  $s_2$ , if there exists a common noun, then the two sentences have noun overlap. If there are two nouns (one from  $s_1$  and the other from  $s_2$ ) sharing a common stem, then the two sentences have argument overlap” (Graesser et al., 2004, p. 199). If a noun from  $s_1$  has a stem that is shared by any category of word in another sentence, then the two sentences have stem overlap. If two sentences share a content word, then they have content word overlap. Lexical overlap contributes to comprehensibility and coherence of the text (Crossley, Weston, Sullivan, & McNamara, 2011) and therefore, was measured in the present study.

*Semantic coreferentiality.* This is measured in Coh-Metrix with Latent Semantic Analysis (LSA), a mathematical technique for representing deeper world knowledge based on large corpora of texts. “LSA uses a general form of factor analysis to condense a large corpus of texts down to 300x500 dimensions. These dimensions represent how often a word occurs within a document (defined at the sentence level, the paragraph level, or in larger sections of texts) and each word, sentence, or text is represented by a weighted vector. The relationships between the vectors form the basis for representing semantic similarity between words” (ibid, p.292). Semantic coreferentiality is an important indicator of cohesion and coherence (ibid).

*Anaphoric reference.* This aspect is measured in Coh-Metrix by comparing pronouns to previous references. For a pronoun in a sentence, “Coh-Metrix calculates if previous sentences contain noun or pronoun references that agree in number (singular/plural), gender (male/female), and person (human/nonhuman) with the given pronoun. Anaphoric references are important indicators of text cohesion” (ibid, p.292).

In Coh-Metrix, a distinction is made between local and global cohesion and coherence (Givon, 1993; Louwrese, 2002; McNamara et al., 1996; Van Dijk & Kintsch, 1983, cited in McNamara et al., 2002). Local cohesion and coherence (Coh\_Local) indicates relations between adjacent clauses and is measured by incidence of connective, content word overlap, adjacent argument overlap, adjacent stem overlap, adjacent

anaphor reference, LSA sentence adjacent. Global cohesion and coherence (Coh\_Global) indicates relations between groups of clauses and groups of paragraphs and is measured by argument overlap, stem overlap, anaphor reference, LSA sentence all, LSA paragraph. The present study used both local measures and global measures for cohesion and coherence.

It is acknowledged that Coh-Metrix is better for analyzing cohesion than for coherence since how coherent a text is from the reader's perspective cannot be measured. However, it is a useful tool to complement subjective ratings on texts.

A total of eighteen Coh-Metrix indices belonging to the categories of measures described in previous paragraphs were used (see Table 5). Following a study that investigated cohesion and coherence of texts written by Chinese ESL learners (Liang, 2006), principal component factor analyses were carried out using SPSS, because there were multi-collinear relationships between these variables and for the convenience of reporting the results. Two factors were extracted, one for local cohesion and coherence and one for global cohesion and coherence. The results of KMO tests were greater than .6 and the Bartlett's tests of sphericity were all significant ( $p < .001$ ), indicating that factor analysis was appropriate. The values of these two factors will be used in reporting the results for organizational features of the essays, thus negative figures may appear.

Table 5 Summary of Coh-Metrix variables included in the present study

Variables		Descriptions and Examples
Coh_Local	1. All connectives	Incidence of all connectives e.g. if, and but, unless
	Temporal connectives 2. Positive 3. Negative	Incidence of temporal connectives After, before until
	Causal connectives 4. Positive 5. Negative	Incidence of causal connectives Because, so nevertheless
	Additive connectives 6. Positive 7. Negative	Incidence of additive connectives And, moreover But, however
	8. Conditional operators	Incidence of conditional expressions e.g. if, in case
	9. Content word overlap	Proportion of content words that overlap between adjacent sentences
	10. Adjacent argument overlap	Overlapping noun, pronoun, or noun phrase (NP) in adjacent sentences
	11. Adjacent stem overlap	Referential overlap between a noun, pronoun or NP and a proposition that has a similar morphological stem in adjacent sentences
	12. Adjacent anaphor reference	Anaphoric reference in adjacent sentences
	13. LSA sentence adjacent	Similarity between adjacent pairs of sentences
Coh_Global	14. Argument overlap	Overlapping noun, pronoun, or NP in all possible pairs of sentences
	15. Stem overlap	Referential overlap between a noun, pronoun or NP and a proposition that has a similar morphological stem in all possible pairs of sentences
	16. Anaphor reference	Anaphoric reference in all possible pairs of sentences
	17. LSA sentence all	Similarity between all possible pairs of sentences
	18. LSA paragraph	Similarity between a paragraph and the other paragraphs

In sum, the organizational features of the essays were analyzed by:

1. Subjective rating by two experienced ESL teachers using an IELTS writing band

descriptor as the marking criteria.

2. Automated scores on coherence and cohesion yielded by Coh-Metrix.

#### **4.8.2 Questionnaires and interviews**

Three pen and paper questionnaires were used in this study. The first one required demographic information from the participants. Answers to the questions provided factual information and were sorted according to the questions in the questionnaire (see Appendix B). In the post-task questionnaire and the post-study questionnaire, there were open-ended questions and closed ended questions (see Appendix E and Appendix G). Answers to closed ended questions, which used a five-point Likert scale, were directly submitted to statistical analysis. The answers to open-ended questions were analyzed qualitatively. The method for coding this part of the data is the same as that for the interview data, which will be described in the following paragraphs.

To complement data from the questionnaires, a semi-structured interview was conducted, in which half of the participants (see section 4.7.6 for selection procedure) were interviewed to answer questions on the process of planning and writing. The questions for the pre-task planners included what they did during planning, what their planning focus was, how they prepared for content and language, and how much of the plan was applied in the subsequent writing, etc. The questions for the online planners and the non-planners concerned their process of writing, while-writing focus, and comments on their task completion conditions (see Appendix F for interview questions). Answers to the interview questions were transcribed verbatim and printed out for analysis. A second rater, who was an experienced ESL instructor and qualitative researcher in a Beijing university, was asked to analyze the qualitative data. In the first step, the second rater was given part of the transcription and the researcher and the second rater went through the data individually, conducting a content analysis<sup>9</sup> of the answers. The answers to

---

<sup>9</sup>Content analysis is defined as “a research technique for making replicable and valid inference from texts to the contexts of their use” (Krippendorff, 2005). In this study the term is used to refer to

interview questions and the questions in the questionnaire were categorized into themes and patterns. For questions regarding planning strategies and writing strategies, the directed approach<sup>10</sup> (Hsieh & Shannon, 2005) was adopted in the analysis. Categories from Ortega (2005) and Guar-Tavares (2008)'s study were used as reference. When there was disagreement on classifications of strategies, the researcher and the second rater discussed the samples of answers in order to reach consensus. The following excerpts are examples of planning strategies.

S3<sup>11</sup>: "Sometimes I really wanted to use a good word but I forgot (the form), then I would use a very common word instead." (lexical compensation)

S14: "After I read the points on the task sheet I thought of the sentence patterns and phrases or good words I may need to write about that." (lexical and syntactical search)

In the end, we worked together to classify L2 planning strategies and writing strategies, which will be fully reported in Chapter 7.

With respect to questions regarding the focus of planning and writing and comments on task completion conditions, a conventional approach of content analysis was adopted. With this approach, "coding categories are directly derived from text data" (Hsieh & Shannon, 2005, p.1). For this part of the data, the researcher first went through one third of the qualitative data, identified themes and patterns, and developed a preliminary coding scheme. Then the coding scheme and part of the data were given to the second rater to code independently. After that, the researcher and the second rater met to compare their coding and the coding scheme was finalized. The full lists of categories for the focus of planning, planning strategies, strategies used during writing, and learners' views on their respective planning condition are presented in Chapter 7. After the coding scheme was finalized, the researcher and the second rater coded the whole set of data independently.

---

systematic analysis of the content of open-ended questionnaire data and transcript of interviews through a process of classification and identification of the themes or patterns.

<sup>10</sup> In a directed approach, a theory or previous relevant research findings are used as guidance in initial coding.

<sup>11</sup> 'S' stands for 'student' and the following number is the number assigned to the student.



### 4.8.3 Planning notes

The questions regarding how much of the participants' plan was used in writing were answered by comparing their planning notes with their written products. To answer this question, planning notes were analyzed in terms of: 1) the total number of words written on the planning note, and 2) the percentage of the words in the planning note that were applied in the subsequent writing. The question about their planning styles (i.e. way of planning) was answered by first categorizing planning notes and then examining the use of each category.

Measure 2) was analyzed by using Text\_Lex Compare, an online program from Lextutor.ca. With this tool, *the percentage of words in the planning note that were applied* was calculated by dividing the number of words in the planning note that reappeared in the essay by the total number of words in the planning note. This gave some idea of how much of the plan was applied in the subsequent writing. *The style of planning* was analyzed with a method adapted from Risemberg's (1996) organizing/transforming pre-writing scale. The planning styles were categorized into four types: simple listing of words and phrases (Type A), drafting, which means the learners wrote complete sentences like they were drafting the essay (Type B), making connections between ideas by drawing lines between phrases or sentences or grouping ideas (Type C), and a mixture of Type A and B (Type A/B). Type A, A/B, and B were assigned to each planning note by looking at the amount of phrases written down and the amount of complete sentences written down on the planning sheets. Two arbitrary cut-off points were used. If 50 % or more of the written notes were words and phrases, the plan would be assigned to Type A. If words and phrases accounted for less than 10% of the written notes, the plan would be assigned to Type B. A Type A/B plan would have 10%-50% of words and phrases on the written notes. A Type C plan would have to be identified by looking to see if there were lines or arrows between words, phrases or sentences, or if the phrases and/or sentences were grouped according to the different aspects to be compared in the essay. Table 6 below presents examples of the four types of plan, using data taken

from the present study.

Table 6 Examples of four types of plan

Plan type	Example
Type A	not only...but also as far as I'm concerned if i.... compared with which one.... Were I....I prefer Dave More free time Young                      Jac: old Traveled.... More experience
Type A/B	Dave Young energetic His education background is more compatible with travel agent Out-going, quick-tempered social, easier to attract more guests Won't spend too much time on children and family
Type B	I prefer that.... Here I'll state my reasons on this matter. You can see the A1 is closer to city center. But A2 is more convenient to Bus station, change to take subway. Meanwhile, the price of per square is much cheaper. Taken the size into account, we still save a lot if we chose A2. That can make up for the extra cost in traffic. For the aspect of level of amenity and facilities, A2 is more likely to provide the chance to build up our bodies. And scenery in a suburb will erect a well-designed nature surrounding. It will take ...40 hrs, but we gain more on other of the.. Last but not least, which I mean, quite important is the neighborhood
Type C	baby: boys always need more attention-----too naughty, girls are clean. pupil: boys need more time and energy to take care of, girls are caring and considerate. Teenager: boys like doing against parents' will---hard to control, girls are sensitive and can take care of parents. Adult: boys need more money---have to marry a girl, buy house, arrange marrying ceremony, etc.

## 4.9 Data Analysis

### 4.9.1 Analysis of learners' writing performance

Essays written in the treatment sessions and the three testing sessions were analyzed in the following steps. The first step was to carry out descriptive analyses to get an

overview of the dependent variables (i.e. accuracy, complexity, fluency, and organization) of written texts. Descriptive statistics provide minimum, maximum, and the mean values of the results for each of the measures described in section 4.8, as well as the standard deviations of each group.

The next step was to check the normal distribution of data on all dependent variables by doing the Kolmogorov-Smirnov (K-S) test on SPSS. For data, which were not normally distributed, transformation of data was conducted by using the square root function on SPSS. If it became normally distributed after transformation, parametric tests were applied in statistical analysis, using the transformed values. If data was not normally distributed even after transformation, non-parametric tests were used in later analysis.

In parametric tests, a one-way ANOVA was used to check if there were between-group differences in the pre-test. Repeated measures ANOVAs were run to find out the significance of differences between and within the three groups in the four treatment sessions and across the three tests. In non-parametric tests, Kruskal-Wallis tests were used to detect if there were any significant between-group differences in the three tests or in the treatment sessions. Friedman's ANOVAs were used to examine the within-group differences across time.

#### **4.9.2 Analysis of planning notes**

Repeated measures ANOVAs were performed to detect the differences among (1) the total number of words jotted down on the planning sheets and (2) the mean percentages of words on the planning sheets that were applied. Repeated measures ANOVA tests were selected for the reason that the data was from the same group of participants, who performed four tasks at different times. This was the appropriate test for analysis of changes in participants' plan applications over time (Field, 2005). Regarding the styles of planning, four types of planning were assigned to a total of 25 pre-task plans for each treatment session, rendering the numbers of each type too small to be statistically analyzed, thus raw numbers were used in reporting the results.

### **4.9.3 Analysis of questionnaire and interview data**

As has been described in the previous section, a content analysis was conducted for answers to open-ended questions in the questionnaire and interviews. At first, the focus was on ideas expressed in the answers. Then themes and patterns were classified into categories. The coded categories were counted and the raw number of times each category was reported and the percentages of participants reporting each category will be presented in Chapter 7.

### **4.9.4 Inter-rater reliability analysis**

Three raters analyzed different portions of the data in order to establish reliability estimates. The first rater was a native English speaker, who holds a PhD in education and used to work as an MA supervisor in a college in New Zealand. She was given 20% of the written data for analysis of accuracy. The identified errors marked by her and the ones identified by the researcher were compared. The percentage of agreement reached 92%.

The second rater was a PhD student in the Department of Applied Language Studies and Linguistics of University of Auckland who has been an ESL instructor in a Chinese university for over ten years. 20% of the written data was given to her for analysis of syntactic complexity. The researcher first explained to her the definitions of T-unit, clause, and the criteria for making judgments. Next, she and the researcher worked independently to identify T-units and clauses. Then Pearson Product Moment Correlation was performed to check the strength of the relationship between the analyses of T-units and clauses. The correlations between the identified T-units and clauses were positive and significant ( $r=.910$ ,  $p<.01$ ).

The third rater, who is an experienced qualitative study researcher and has been an ELS instructor in a university in Beijing for 15 years, was given the whole set of interview data. She was given the coding scheme and then coded the data independently. The agreement between this rater and the researcher was 97%.

An inter-rater reliability analysis was also conducted for the subjective rating for the

organization of the written texts. As has been explained, two raters were recruited for this job. The detailed rating procedure is as follows. The raters were presented with the band descriptor and studied the criteria for scoring. Then they were given 10 essays randomly selected from the data pool to rate independently. The marks they awarded to each of the essays were compared and discussed in relation to the band descriptor. After discussion, consensus was reached in regards to the understanding of the criteria and the appropriate marks to award to the essays. Then the raters took all the essays home to rate independently.

The interrater reliability between the two raters was calculated with Pearson Product Moment correlation coefficient test and the correlation was not high ( $r=0.69$ ). This was not surprising considering the nature of this type of rating. Studies on marking reliability showed that marking reliability is dependent on the subject area (C. James, 1974; McVey, 1975) and the type of scripts being assessed (Byrne, 1979). C. James (1974) and McVey's study (1975) revealed that marking reliability in examinations of scripts in physics and electronic engineering was high (correlations were 0.9 or above). Byrne (1979) found that interrater reliability was highest for assignments in mathematics and physics and low in arts and social sciences. In terms of the type of scripts being assessed, essay questions presented the greatest reliability problem, irrespective of subject area. Considering the subjective nature of essay rating in the present study and the inadequate training and experience in using the IELTS writing band descriptors (though the two raters were experienced in teaching ESL writing), the low interrater reliability was not surprising.

To remedy the low reliability, the following method was used. For the scores that were in exact agreement (two raters gave the same scores to one essay) and the adjacent scores (two scores given by the two raters differed by 1 point), the average scores of the two raters were used as the operational scores. For the scores given by the two raters that differed by 2 points, the essays were rescored by a third rater (the researcher of the present study). In these cases, the average scores of the three raters were used in later statistical analysis. The remedy method was appropriate because studies have shown that it can successfully increase marking reliability (R. L. Johnson, Penny, & Gordon, 2000; R.

L. Johnson, Penny, & Gordon, 2001) of essay ratings.

#### **4. 10 Pilot study**

A pilot study was carried out from April 29<sup>th</sup> to May 17<sup>th</sup>, 2009 to try out the instruments, testing procedures, and experimental procedures. At that time, the study was designed to look into the effects of pre-task planning only; therefore, participants were put into two groups: a pre-task planning group and a no planning group. The pre-task planning group had 10 minutes to plan prior to finishing the writing task in 15 minutes. In total, the pre-task planning group had 25 minutes to complete a writing task. The control group was also given 25 minutes in total to ensure that it had the same amount of total on task time. In the first 15 minutes, they carried out the same writing task as the one completed by the pre-task planning group under a no planning condition. Then they were given another task similar to the task they had just completed but different in topic to finish in 10 minutes.

The participants were given a pre-test and a post-test (see section 4.6.1 for detail description). Prior to the pre-test, they were given a questionnaire to fill out to provide demographic information. After the pre-test, there were three experimental sessions, in which the writing tasks were tried out. Following each experimental session, a post-task questionnaire was given to gather information on what the learners did during planning (see section 4.6.2 for details). This was followed by the semi-structured interview (see section 4.6.3) designed to provide more in-depth information on learners' process of planning and task completion process. Then a post-test was administered followed by a post-study interview to find out learners' attitudes towards writing in English in general, their feelings about making mistakes, how much they knew about the target structure, and whether they had realized which linguistic feature was targeted. An overview of the pilot study design is displayed in the figure below.

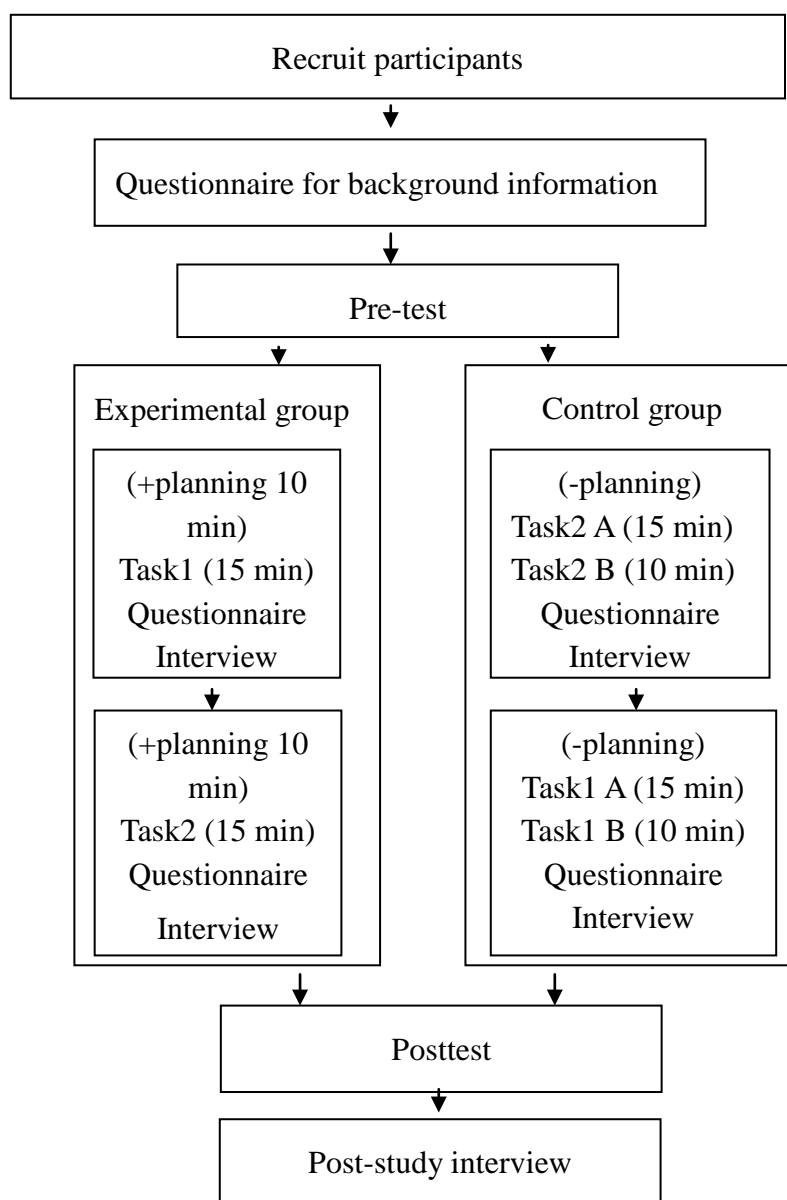


Figure 3 Overview of the pilot study

The participants of the pilot study were seven students from a language school in Auckland. They were all from Academic English Module 2 classes, and were categorized as intermediate level learners. At the time data was collected, they had had classes in Module 2 for one week. Four of them proceeded into Module 2 classes from Module 1. Prior to Module 1, they had had one year of study in the General English Module. The rest of the participants were placed in Module 2 classes according to their scores in a placement test. Among the seven participants, four were from Saudi Arabia, one from Romania, one from Chile, and one from Kazakhstan. Their years of learning English

ranged from five months to seven and a half years. There were three females and four males, and the age range was from 20 to 37. From the information above, we could see that they were different in L1, age, and English learning background but similar in English proficiency level.

Eight writing tasks were designed for the pre-test, post-test and the experimental sessions (see section 4.6.1 for detail description). Both the experimental group and the control group did Task 1, Task 2, and Task 3. The control group also did Task 1B, Task 2B, and Task 3B. They were all letter writing tasks requiring learners to persuade their families or friends to agree with their decisions. Two of the tasks were used in the pre-test and the post-test and the other six were used in the experimental sessions.

The writing tasks were measured in terms of accuracy of comparative forms, the overall accuracy, complexity, and fluency. Since the sample size was very small, statistical analysis would not be significant; thus the results are not reported.

#### Limitations of the pilot study and its implications for the main study

1. From the feedback of the participants one of the writing tasks was distinctively easier than other tasks. It did not generate sufficient amount of written output or number of occurrences of comparative forms. For these reasons, the task was discarded from the task pool.
2. The accuracy rate of comparative forms used in written data was relatively high, causing a concern for the appropriateness of using this feature as the target structure. However, the results were from a very small sample size so they were not convincing enough for a decision to discard this target structure to be made. Therefore, it was decided that S-V agreement should be added as another target structure to be examined in the main study.
3. It was found that the design of the pilot study had a potential problem, in that the control group had more practice than the experimental group, which might lead to better performance of this group in the post-test. To avoid confounding the effects of practice with the effects of planning, the main study employed the following



design. Instead of dividing participants into two groups, the participants were divided into three groups. PTP was the pre-writing planning + limited on-line planning group; OLP was the no pre-writing planning + more on-line planning group; and the NP was the no pre-writing planning + limited on-line planning group. In one treatment, participants in one group would do one task only.

4. In the pilot study, only a pre-test and a post-test were administered because of the time constraint. That is, only the short-term effects of planning on L2 writing performance were investigated. The purpose of the main study was to examine the relationship between planning and L2 development as well as the immediate effects on L2 writing, a delayed post-test was added in the main study.
5. In the pilot study, the post-task questionnaire and the interview were conducted in English. Because of the proficiency level of the participants, they had some difficulty understanding the questions and expressing themselves. That caused the loss of some information and the data were not as rich as expected. Therefore, the post-task questionnaire and the interviews were conducted in Chinese in the main study. This helped overcome the language barrier in communication and yielded richer data.
6. The timing of the post-task questionnaire and interview was changed in the main study for the sake of preventing the possible reactive effects of giving retrospective verbal reports on the subsequent task performance. Instead of asking all participants to fill out a questionnaire after each task, they were given the questionnaire after the immediate post-test to provide information on their planning and task completion processes. After completing the questionnaire, half of the participants were interviewed one-on-one to provide more in-depth information on how they had planned and completed the tasks.
7. Because the main study had a much larger sample size than the pilot study, it was difficult to conduct a one-on-one post-study interview. Therefore, a post-study questionnaire was used instead. After the delayed posttest, all participants were asked to fill in the post-study questionnaire to answer questions such as whether

they had grammar instructions during the period of the study, what they thought the focus of the study was, their general attitudes on writing in English, and about making mistakes, etc.



## **CHAPTER 5 THE EFFECTS OF TASK PLANNING ON IMMEDIATE L2 WRITING PERFORMANCE**

### **5.1 Overview**

The research question this chapter addresses is: Does task planning have an effect on immediate L2 writing performance?

In this study, participants were put into three treatment groups: no-planning, pre-task planning, and online planning group. They completed writing tasks under these different experimental conditions once a week for four weeks. In order to answer the research question, the three groups' writing performances in the treatment sessions were compared and the within-group differences were also looked at.

In the following sections, an overview of data analysis will be presented first. This will include a summary of dependent variables, assumption checking for parametric tests (i.e. normality of data distribution), data transformation, and calculation of effect sizes. This will be followed by the results of the statistical analysis of the independent and dependent variables. After the presentation of results, this chapter will conclude with a discussion.

### **5.2 Overview of data analysis**

Following previous studies on task planning and informed by the literature on performance measurement, the present study used 12 dependent variables to measure learners' writing performance. The table below summarizes the dependent variables. For more information about and description of the dependent variables see Chapter 4.

Table 7 Summary of dependent variables

Dependent variables	Label		Measure	
Accuracy Specific measures	1	Target structure: Comparative	1	Accuracy of comparative forms
	2	Target structure: S-V	2	Accuracy of S-V agreement
General measures	3	%error-free	3	Percentage of error-free clauses
	4	Error/100	4	Errors per 100 words
Syntactic complexity	5	MLT	5	Mean length of T-units
	6	C/T	6	Number of clauses per T-unit
Lexical complexity	7	D	7	Type-token ratio <sup>12</sup> (lexical diversity)
	8	Lam	8	Percentage of advanced words <sup>13</sup> (lexical sophistication)
Fluency	9	WPM	9	Words per minute
	10	Dysfluency	10	Ratio of self-corrections
Organization	11	Organization rating	11	Scores of coherence and cohesion given by raters
	12	Coh_Local	12	Scores of local coherence and cohesion generated by Coh-Metrix
	13	Coh_Global	13	Scores of global coherence and cohesion generated by Coh-Metrix

Preliminary to an analysis of the dependent variables listed in the table above, all data were graphed and plotted in SPSS 16.0 for frequency distribution. Boxplots were checked to pinpoint the outliers. Normal Q\_Q plot and Kolmogorov-Smirnov (K-S) tests were applied to check the normality of distribution of dependent variables for each group. Following the results of K-S tests, all variables that deviated from normality ( $p < .05$ ) were transformed using a square root transformation. After transformation, measures of the two target structures, the mean length of T-units, and the subjective rating on coherence and cohesion were still not normally distributed. Thus these were subjected to non-parametric tests. Normally distributed data was analyzed with repeated measures

<sup>12</sup>Adjusted total number of different words divided by total number of words.

<sup>13</sup>The number of advanced words divided by the total number of words.

ANOVAs, testing for simple group effects, simple time effects, and interaction between group and time. Since the groups significantly differed on the pre-test for lexical diversity (see Chapter 6), repeated measures ANCOVA test was used for this measure.

With parametric and non-parametric tests, statistical significance could be checked to see if the differences, if any, were caused by chance. However, these tests are affected greatly by sample size and the magnitude of an effect cannot be displayed. For these reasons, effect sizes were also calculated since they could provide an objective measure of the importance of an effect. Two commonly used effect size measures, *r* and *d*, were used in the present study. Cohen's *d* is the most commonly used measure in the field of applied linguistics and its calculation is based on the means, standard deviations, and sample sizes of the treatment and control groups in an experimental study. It is used for calculating effect sizes for differences between two groups. Cohen's *d* was calculated when post hoc group contrasts were conducted after a significant main effect was detected. An effect size is considered small when *d* is .20; it is medium when *d* is equal to .50; and it is large when *d* is equal to or larger than .80. For non-parametric tests, *r* was used to measure effect size using the formula suggested by Field (Field, 2005), which is as follows:  $r = \frac{Z}{\sqrt{N}}$ <sup>14</sup>. An effect size is considered small when *r* is .10; it is medium when *r* reaches .30 and large when *r* is equal to or exceeds .50. Apart from *d* and *r*, Cohen's *f* is used for calculating the magnitude of the main group effect in preliminary analyses (Repeated ANOVAs) because it is an appropriate measure for differences among three or more groups (Kinnear & Gray, 2009). Following effect size conventions, *f* scores of .10, .25, and .40 are considered small, medium and large respectively (Cohen, 1992). For clarity, Table 8 below shows how the effect sizes were interpreted.

---

<sup>14</sup>*Z* is the *Z* score produced by SPSS; *N* is the number of total observations on which *Z* is based. This formula was originally provided by Rosenthal (1991) and cited in Field (2005).

Table 8 Interpretation of effect sizes

	Cohen's d	Cohen's f	r
Negligible	$d < .20$	$f < .10$	$r < .10$
Small	$.20 \leq d \leq .49$	$.10 \leq f \leq .24$	$.10 \leq r \leq .29$
Medium	$.50 \leq d \leq .79$	$.25 \leq f \leq .39$	$.30 \leq r \leq .49$
Large	$d \geq .80$	$f \geq .40$	$r \geq .50$

### 5.3 Results for the independent variable

As has been described in Chapter 4, three task conditions were established: no planning which had 15 minutes to complete the writing tasks without any pre-task planning time, pre-task planning which had 10 minutes to prepare before the tasks and 15 minutes to write, and online planning which had 25 minutes to write. While it was intended that the online planning group should spend more time writing the essays than the other groups, it is important to ascertain whether the participants performed the tasks as required. Since there was no record of what each online planner was thinking and doing during task time (except their self-report data), there was no hard evidence that the online planning group did engage in online planning more than the other two groups. However, the total number of words produced in each task was some evidence to show that the OLP spent the extra time on writing rather than doing other things. As is shown in the table below, the online planning group (OLP) produced the biggest number of words and ANOVA tests show that the between-group difference was statistically significant ( $F=4.819$ ,  $p=.010$ ). Post hoc comparisons show that the differences between OLP and both PTP ( $p=.020$ ) and NP ( $p=.025$ ) were significant. This is some evidence that the treatments were operationalized as intended.

Table 9 Results for the total number of words produced in the experimental sessions

Measure	Means			F-value	Sig.	Group contrasts		
	PTP	OLP	NP			PTP-NP	PTP-OLP	NP-OLP
Total words	186.04	213.2	184.8	4.819*	<b>.010</b>	.992	<b>.020</b>	<b>.025</b>

### 5.4 Results for the dependent variables

In the following sections, results for accuracy, complexity, fluency, and organization

are reported separately. All the sections will be organized in the same way: 1) descriptive statistics of the measures; 2) inferential statistics and effect sizes; 3) results of post hoc comparisons when preliminary tests indicated a significant difference.

#### 5.4.1 Results for accuracy measures

In the present study, the accuracy of students' writing was measured by the percentage of accurate use of comparative forms (comparative) and S-V agreement (S-V), the percentage of error-free clauses (%error-free), and the number of errors per 100 words (error/100). Table 10 below displays the descriptive statistics of these measures. Information on obligatory occasions for suppliance of the two target structures (Comparative and S-V) is also provided in the table, since results for accuracy of the target structures should be interpreted in relation to this information. That is, the more frequently a form is used the greater the chances for errors to be made.

Table 10 Descriptive statistics for the accuracy measures in the treatment sessions

Measures	Group	N	T1			T2			T3			T4		
			Obl (M)	M	SD	Obl (M)	M	SD	Obl (M)	M	SD	Obl (M)	M	SD
Comparative	PTP	25	5.84	.90	.45	5.28	.92	.16	6.36	.94	.13	6.24	.94	.12
	OLP	25	6.52	.94	.12	6.16	.98	.06	5.56	.96	.09	4.96	.94	.13
	NP	25	5.96	.93	.18	5.32	.98	.06	5.16	.92	.17	5.72	.90	.15
S-V	PTP	25	16.8	.95	.07	16.0	.98	.04	16.2	.96	.05	16.9	.94	.10
	OLP	25	19.4	.91	.10	18.0	.94	.07	17.2	.93	.09	17.2	.95	.06
	NP	25	18.6	.94	.07	17.0	.96	.06	15.8	.94	.07	15.8	.96	.06
%error-free	PTP	25		.72	.12		.67	.10		.69	.13		.66	.13
	OLP	25		.64	.14		.68	.17		.68	.15		.68	.12
	NP	25		.68	.17		.70	.11		.67	.12		.71	.14
Error/100	PTP	25		.05	.02		.06	.02		.06	.02		.06	.02
	OLP	25		.07	.02		.06	.03		.06	.03		.06	.02
	NP	25		.06	.03		.06	.02		.06	.02		.05	.03

Notes: 1. T1=Time 1; T2=Time 2; T3=Time 3; T4=Time 4; 2. Obl(M)=mean of obligatory occasions for suppliance of the form

The measures for the two target structures violated the assumption of normality distribution for parametric tests. Therefore, Kruskal-Wallis tests were run to examine the between-group differences and Friedman's ANOVA tests were run to investigate the



within-group differences. The results of these tests are summarized in Table 11. The two general measures for accuracy were normally distributed, thus were submitted to repeated measures ANOVAs for a check of significant differences, the results of which are displayed in Table 12.

Table 11 Results for the target structures in the treatment sessions

Kruskal-Wallis	Measures	Chi-square	df	Asymp. Sig.	Effect sizes (r)
	T1 Comparative	2.961	2	.228	.09
	S-V	3.497	2	.174	.24
	T2 Comparative	4.406	2	.110	.18
	S-V	4.675	2	.097	.27
	T3 Comparative	.803	2	.669	.13
	S-V	.428	2	.807	.18
	T4 Comparative	2.588	2	.274	.16
	S-V	1.104	2	.576	.08
Friedman's ANOVA	Comparative	4.895	3	.180	.08
	S-V	2.638	3	.451	.15

Table 12 Results for the general measures for accuracy in the treatment sessions

Measures	Main effects	Df1	Df2	F	Sig.	Effect sizes (f)
Error/100	Within-group	3	216	.331	.803	.07
	timexgroup	6		1.281	.267	.19
	Between-group	2	72	.343	.711	.10
%error-free	Within-group	3	216	.085	.968	.03
	timexgroup	6		1.363	.231	.19
	Between-group	2	72	.284	.754	.09

It can be seen from Table 11 that there were no significant differences between groups in the use of the two target structures in any of the four treatment sessions. The effect sizes for comparative forms were small in all sessions. The effect sizes for S-V agreement were small in Time 1 (T1), Time 3 (T3), Time 4 (T4), and medium in Time 2 (T2). The within-group differences were not significant either ( $\chi^2(3)=4.895$ ,  $p=.180$ ;  $\chi^2(3)=2.638$ ,  $p=.451$ ) and the effect sizes were small.

With respect to the general measures for accuracy, results show that differences in neither of the measures were statistically significant. For *errors per 100 words*, planning did not show any significant effect as seen in between-group difference ( $F(2, 72)=.343$ ,  $p=.711$ ). There were also no within-group differences either ( $F(3, 216)=.331$ ,  $p=.803$ ). In

terms of the *percentage of error-free clauses*, the between-group effects were not significant ( $F(2, 72)=.284, p=.754$ ). Learners did not change significantly across tasks either as shown in the non-significant within-group difference ( $F(3, 216)=.085, p=.968$ ). The effect sizes for between-group differences and within-group differences were negligible, as they were all smaller than .10, the threshold for a small effect size.

#### 5.4.2 Results for syntactic complexity measures

Two measures for syntactic complexity were employed in the present study: *mean length of T-units* (MLT) and *number of clauses per T-unit* (C/T). Table 13 below displays descriptive statistics for these measures.

Table 13 Descriptive statistics for syntactic complexity in the treatment sessions

Measures	Group	N	T1		T2		T3		T4	
			M	SD	M	SD	M	SD	M	SD
MLT	PTP	25	11.77	2.49	11.46	2.38	11.56	2.31	11.03	2.0
	NP	25	11.60	2.63	11.58	1.98	11.55	2.07	12.21	3.12
	OLP	25	11.64	1.91	12.01	2.71	12.96	3.07	11.65	1.62
C/T	PTP	25	1.80	.33	1.79	.28	1.77	.29	1.66	.30
	OLP	25	1.79	.31	1.73	.29	1.76	.28	1.80	.38
	NP	25	1.84	.40	1.87	.44	1.86	.34	1.78	.30

Table 14 Results for syntactic complexity measures in the treatment sessions

Kruskal-Wallis	Measures	Chi-square	df	Asymp. Sig.	Effect sizes (r)
	T1 MLT	.605	2	.739	.04
	T2 MLT	.340	2	.844	.09
	T3 MLT	3.423	2	.181	.22
	T4 MLT	2.714	2	.257	.30
Friedman's ANOVA	MLT	4.488	3	.213	.10
Repeated measures ANOVA	C/T	F	df	Sig.	Effect Size (f)
Between-group		.938	2(72)	.396	.16
Within-group		.836	3(162)	.476	.11

The data for *mean length of T-units* (MLT) was not normally distributed even after it was transformed, thus was submitted to non-parametric tests. The data for *number of*

*clauses per T-unit* (C/T) met the normal distribution standard as tested by K-S, therefore, repeated measures ANOVA were used. The results of these tests are reported in Table 14.

Results show that planning did not have significant effects on syntactic complexity as measured by MLT, as the scores of the three groups did not differ significantly from one another in the four sessions ( $H(2)=.605$ ,  $p=.739$ ;  $H(2)=.340$ ,  $p=.844$ ;  $H(2)=3.423$ ,  $p=.181$ ;  $H(2)=2.714$ ,  $p=.257$ ) and the effect sizes were small in T1, T2, T3, and medium in T4. Learners' performance in MLT did not significantly change over time either, for the within-group difference was not significant ( $\chi^2(3)=4.488$ ,  $p=.213$ ) with a small effect size ( $f=.10$ ). With regard to C/T, planning did not show any effect as the between-group difference was not significant ( $F(2, 72)=.938$ ,  $p=.396$ ). Time did not have a significant effect either ( $F(3, 162)=.836$ ,  $p=.476$ ). The effect sizes were small ( $f=.11$ ;  $f=.16$ ).

### 5.4.3 Results for lexical complexity measures

In the present study, lexical complexity was measured by *lexical diversity* (D) and *lexical sophistication* (Lam). The descriptive statistics of these measures are reported in Table 15 below.

Table 15 Descriptive statistics for lexical complexity in the experimental sessions

Measures	Group	N	T1		T2		T3		T4	
			M	SD	M	SD	M	SD	M	SD
D	PTP	25	80.48	21.80	86.99	25.18	80.17	17.14	80.83	18.55
	OLP	25	83.26	23.07	80.83	21.53	74.84	18.83	80.74	29.42
	NP	25	76.16	21.52	77.82	15.09	71.92	18.0	83.83	27.42
Lam	PTP	25	1.04	.38	1.19	.33	1.28	.45	1.01	.36
	OLP	25	.93	.28	1.06	.31	1.03	.33	1.00	.42
	NP	25	.92	.44	1.00	.38	1.05	.50	.94	.35

Data on lexical diversity (D) was tested by ANCOVA because a significant between-group difference was found in the pre-test (see Chapter 6 Section 6.2.1 for details). Data on lexical sophistication was analyzed with repeated measures ANOVA since the assumption of normal distribution was tenable in this measure. The results of these tests are displayed in Table 16.

Table 16 Results for lexical complexity in the treatment sessions

Measures	Main effects	Df1	Df2	F	Sig.	Effect sizes(f)
D	Within-group	2.51	180.571	1.70	.168	.15
	timexgroup	5.02		.736	.621	.14
	Between-group	2	72	.705	.497	.14
Lam	Within-group	2.67	191.834	2.203	.097	.18
	timexgroup	5.33		.216	.962	.08
	Between-group	2	72	2.972*	.050	.29

Table 17 Pairwise comparison results of lexical sophistication (Lam) in the treatment sessions

Group I—Group J	Mean Difference (I-J)	Std. Error	Sig.	Effect size (d)
NP----OLP	-.028	.056	.621	.08
NP----PTP	-.129*	.056	<b>.023</b>	.37
OLP---PTP	-.101	.056	.073	.33

Results show that learners' performances in *lexical diversity* (D) were not significantly affected by planning conditions ( $F(2, 72)=.705$ ,  $p=.497$ ). Learners' performances over time were not significantly different either ( $F(2.51, 180.571)=1.70$ ,  $p=.168$ ). The assumption of sphericity was broken ( $\chi^2(5)=19.788$ ,  $p=.001$ ) so degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ( $\epsilon=.836$ ). The effect sizes were small ( $f=.15$ ,  $f=.14$ ).

As for *lexical sophistication* (Lam), there was some evidence of an effect for planning ( $F(2,72)=2.972$ ,  $p=.05$ ). The effect size for between-group difference was medium ( $f=.29$ ). An examination of descriptive data (Table 15) revealed that the PTP group used more advanced words than the other two groups and the OLP group generally produced more advanced words than the NP group. Pairwise comparison results (Table 17) show that PTP was significantly different from NP ( $p=.023$ ) while OLP was not. The effect sizes for differences between PTP and the other two groups were medium ( $d=.37$ ;  $d=.33$ ). These results suggest that pre-task planning helped learners produce more lexically sophisticated words than those who did not have time to do pre-task planning.

With regard to the change over time in this measure, Table 16 shows that the change was not significant ( $F(2.67, 191.834)=2.203$ ,  $p=.097$ ) and the effect size was negligible.

#### 5.4.4 Results for fluency measures

There were two measures for fluency used in the present study: one was for measuring temporal fluency (WPM) and one was for measuring repair fluency (Dysfluency). Table 18 below presents the descriptive statistics for these two measures. To test if differences between group means were statistically significant, repeated measures ANOVAs were run as the data satisfied the assumption of normal distribution. The results of ANOVA are reported in Table 19.

Table 18 Descriptive statistics for fluency measures in the treatment sessions

Measures	Group	N	T1		T2		T3		T4	
			M	SD	M	SD	M	SD	M	SD
WPM	PTP	25	12.30	2.58	11.15	2.72	12.35	2.66	12.15	2.50
	OLP	25	10.12	2.86	8.71	2.19	8.25	1.67	7.38	1.82
	NP	25	11.92	2.82	12.44	3.08	11.50	2.64	12.10	2.77
Dysfluency	PTP	25	2.91	2.19	3.72	3.38	2.71	2.17	3.13	2.19
	OLP	25	4.64	3.43	4.29	2.90	5.01	3.42	5.22	3.10
	NP	25	4.71	3.86	5.57	7.79	4.78	2.53	6.22	5.03

Note: WPM=number of words per minute; Dysfluency=ratio of number of corrections

Table 19 Results for fluency measures in the treatment sessions

Measures	Main effects	Df1	Df2	F	Sig.	Effect sizes (f)
WPM	Within-group	2.62	188.374	3.45*	<b>.023</b>	.22
	timexgroup	5.23		4.725*	<b>&lt;.001</b>	.36
	Between-group	2	72	23.98*	<b>&lt;.001</b>	.81
Dysfluency	Within-group	1.83	131.743	.885	.450	.11
	timexgroup	3.66		.576	.749	.13
	Between-group	2	72	5.063*	<b>.009</b>	.37

Table 20 Results of post hoc comparisons of WMP in the treatment sessions

Group(I)-----(J)	Mean Difference (I-J)	Std. Error	Sig.	Effect sizes (d)
NP----- OLP	3.37*	.056	<.001	1.40
NP----- PTP	-.0009	.056	1.00	.00
OLP----PTP	-3.37*	.056	<.001	1.47

Table 21 Repeated measures ANOVA results of time effect on WPM by group

Group	SS	Df	MS	F	Sig.	Effect sizes(f)
PTP	23.946	3	7.982	2.455	.07	.32
OLP	98.469	3	32.832	10.834*	<.001	.67
NP	11.329	3	3.776	.924	.434	.20

ANOVA results (Table 19) show that different planning conditions did cause significant between-group differences in temporal fluency ( $F(2,72)=23.98$ ,  $p<.001$ ) and the effect size was large ( $f=.81$ ). Post hoc multiple comparisons (Table 20) reveal that significant differences existed between NP and OLP ( $p<.001$ ) and between PTP and OLP ( $p<.001$ ) and the effect sizes were large ( $d=1.40$ ;  $d=1.47$ ). PTP and NP did not differ significantly from each other ( $p=1.00$ ). The results indicate that the pre-task planning did not lead to a significant advantage in terms of writing speed but online planning had a significantly negative effect on writing speed.

As for the within-group difference, the interaction between time and group was also significant ( $F(5,23)=4.725$ ,  $p<.001$ ,  $f=.36$ ), which shows that the changes of the three groups across sessions were different. A look at the individual performance of each group (Table 21) reveals that the time effect was caused by OLP's decreasing fluency. The time effect on OLP was significant ( $p<.001$ ) while that on PTP ( $p=.07$ ) and on NP ( $p=.434$ ) was not.

Table 22 Results of post hoc comparisons of Dysfluency in the treatment sessions

Group (I)-----(J)	Mean Difference (I-J)	Std. Error	Sig.	Effect sizes (d)
NP----- OLP	.53	.72	.466	.015
NP----- PTP	2.203*	.72	.003	.64
OLP----PTP	1.674*	.72	.023	.78

Regarding dysfluency, the between-group difference again was significant ( $F(2, 72)=5.063$ ,  $p=.009$ ) and the effect size was close to large ( $f=.37$ ). Post hoc multiple

comparisons (Table 22) show that PTP made significantly fewer corrections than NP ( $p=.003$ ,  $d=.64$ ) and OLP ( $p=.023$ ,  $d=.78$ ), while the differences between NP and OLP ( $p=.745$ ) were not significant. It suggests that pre-task planning had positive effects on helping learners make less self-repair in writing.

With regard to the within group difference, learners' change over time was not significant ( $F(1.83, 131.743)=.885$ ,  $p=.450$ ) and the effect size was small ( $f=.11$ ).

#### 5.4.5 Results for organization scores

The written tasks were also assessed in terms of organization of the essays. Three measures were used: subjective rating on text coherence and cohesion (Organization rating) and computer program generated scores on 1) local coherence and cohesion (Coh\_Local); and 2) global coherence and cohesion (Coh\_Global). The descriptive statistics of these measures are presented in Table 23 below. The computer-program-generated scores satisfied the assumption of normal distribution and were analyzed with repeated measures ANOVA. The scores given by raters were not normally distributed and thus were submitted to non-parametric tests. The repeated measures ANOVA results are displayed in Table 24 and the results of the series of non-parametric tests are presented in Table 25.

Table 23 Descriptive statistics for organization measures in the treatment sessions

Measures	Group	N	T1		T2		T3		T4	
			M	SD	M	SD	M	SD	M	SD
Organization rating	PTP	25	5.60	.74	5.54	.68	5.45	.53	5.44	.63
	OLP	25	5.19	.70	5.12	.68	5.15	.74	4.86	.71
	NP	25	5.18	.52	5.21	.65	5.03	.65	4.92	.74
Coh_Local	PTP	25	-.14	.98	-.03	1.04	-.07	.87	-.08	1.02
	OLP	25	-.14	.90	.21	1.08	-.33	.87	.08	.98
	NP	25	.25	1.13	-.13	.86	.39	1.16	-.05	1.05
Coh_Global	PTP	25	.05	.96	-.12	1.01	-.01	.86	.04	.94
	OLP	25	-.23	.91	.16	1.07	-.28	.92	.09	.93
	NP	25	.18	1.16	-.01	.91	.30	1.17	-.14	1.13

Table 24 Results for local and global coherence and cohesion in the treatment sessions

Measures	Main effects	Df1	Df2	F	Sig.	Effect sizes (f)
Coh_Local	Within-group	3	210	.013	.998	.00
	timexgroup	6		1.762	.108	.22
	Between-group	2	72	.707	.496	.14
Coh_Global	Within-group	3	210	.007	.999	.00
	timexgroup	6		1.28	.268	.19
	Between-group	2	72	.425	.655	.11

Table 24 demonstrates that the between-group difference in local coherence was not significant ( $F(2,72)=.707$ ,  $p=.496$ ) in the four treatment sessions and the effect size was small ( $f=.14$ ). The time effect was not significant either ( $F(3,210)=.013$ ,  $p=.998$ ). The same pattern was observed for global coherence. The between-group difference was not significant ( $F(2,72)=.425$ ,  $p=.655$ ) with a small effect size ( $f=.11$ ) nor was the within-group difference ( $F(3,210)=.007$ ,  $p=.999$ ) and the effect size was negligible.

Table 25 Results for subjective ratings on organization in the treatment sessions

Kruskal-Wallis	Measures	Chi-square	df	Asymp. Sig.	Effect sizes (r)
T1	Organization rating	4.996	2	.082	.38
T2	Organization rating	5.41	2	.067	.28
T3	Organization rating	5.164	2	.076	.27
T4	Organization rating	8.783*	2	<b>.012</b>	.35
Friedman's ANOVA	Organization rating	<b>13.320*</b>	3	<b>.004</b>	.27

Table 26 Mann-Whitney results of subjective ratings in Time 4 writing

Groups	Mann-Whitney U	Z	Asymp. Sig.	Effect sizes (r)
NP----- OLP	298.50	-.276	.783	.04
NP ---- PTP	193.50	-2.351	.019	.33
PTP--- OLP	173.50	-2.75*	<b>.006</b>	.39

With respect to the rater given organization scores (Table 25), between-group differences were not significant in T1, T2, and T3 ( $H(2)=4.996$ ,  $p=.082$ ;  $H(2)=5.41$ ,  $p=.067$ ;  $H(2)=5.164$ ,  $p=.076$ ) but were significant in T4 ( $H(2)=8.783$ ,  $p=.012$ ). To investigate this difference further, a series of Mann-Whitney tests were run (Table 26). A Bonferroni correction was applied and the effects were reported at a .0167 ( $0.5/3=.0167$ ) level of significance. The Mann-Whitney test shows that for Time 4 writing, OLP's scores for organization were significantly lower than that of PTP ( $p=.006$ ) with a medium



effect size ( $r=.39$ ). The difference between PTP and NP was not statistically significant but close to significance ( $p=.019$ ) and the effect size was medium ( $r=.33$ ). The difference between OLP and NP was not significant. This indicates that in Time 4 writing, ratings for the organization of texts written by the pre-task planning group were higher than those for the texts written by the other two groups.

With respect to the within-group differences, Table 25 shows that the rater given organization scores changed significantly over time ( $\chi^2(3)=13.32$ ,  $p=.004$ ). To detect further the time effect, a series of Wilcoxon signed ranks tests were conducted. A Bonferroni correction was applied and the effects were reported at a .0083 ( $0.05/6=.0083$ ) level of significance. Results (Table 27) show that learners' organization scores in T4 were significantly lower than those in T1 and T2 and the effect sizes were medium ( $r=.46$ ;  $r=.35$ ).

Table 27 Wilcoxon signed ranks results of subjective ratings in treatment sessions

Tests	Rank	N	Z	Asymp. Sig.	Effect sizes (r)
T2 vs. T1	Negative	28	-.427	.669	.25
	Positive	27			
	Ties	20			
T3 vs. T1	Negative	30	-1.618	.106	.27
	Positive	16			
	Ties	29			
T4 vs. T1	Negative	38	-3.466*	<b>.001</b>	.46
	Positive	16			
	Ties	21			
T3 vs. T2	Negative	31	-1.297	.195	.15
	Positive	21			
	Ties	23			
T4 vs. T2	Negative	36	-3.03*	<b>.002</b>	.35
	Positive	17			
	Ties	22			
T4 vs. T3	Negative	32	-2.028	.043	.25
	Positive	20			
	Ties	23			

## 5.5 Summary

In conclusion, planning had some effects in the following aspects (see Table 28).

Pre-task planning showed an effect in lexical complexity for this group produced significantly more sophisticated words than the no-planning group in the four treatment sessions. There was a negative effect for online planning in temporal fluency, with this group producing significantly fewer words per minute than the other groups. An effect for dysfluency, on the other hand, was evident for pre-task planning, with this group making significantly fewer corrections than the other groups.

As for the within-group differences, there was a significant deterioration over time for temporal fluency for OLP, but there was no significant change for the dysfluency measure. For organization, there was an effect for time for the subjective ratings with performance at T4 weaker than at T1 and T2.

Table 28 Summary of significant between-group and within-group differences

Measures	Time	Group
Accuracy Comparative S-V agreement e/100 %error-free		
Syntactic complexity C/T MLT		
Lexical complexity D Lam		PTP>NP
Fluency WMP Dysfluency	T1>T3, T4 OLP	NP, PTP>OLP PTP<NP, OLP
Organization Organization rating Coh_Local Coh_Global	T1, T2>T4 OLP	PTP>OLP in T4

## 5.6 Discussion

In this section, the results for the immediate effects of planning on L2 written production will be discussed. The effects of planning on the four aspects of language performance (i.e. accuracy, complexity, fluency, organization) will be discussed one after

another. When discussing the effects of planning on each of these performance aspects, the effects of pre-task planning and those of online planning will be discussed separately.

### **5.6.1 Accuracy**

#### **5.6.1.1 Effects of pre-task planning on accuracy**

As has been presented in section 5.4.1, there was no effect on accuracy for pre-task planning in the four experimental sessions. This result is consistent with some previous studies. As has been shown in Chapter 3, findings from previous planning studies regarding the effects of pre-task planning on accuracy have been mixed. In the 26 studies on pre-task planning and oral production reviewed in Chapter 3, eleven did not find any effect on accuracy (e.g. Crookes, 1989; Mehrang & Rahimpoura, 2010; Nakakubo, 2011; Ojima, 2006; Ortega, 1995a; Wendel, 1997). In the other studies, effects on accuracy were found in some measures (e.g. Crookes, 1989; Ortega, 1999; Sangarun, 2001), some tasks (e.g. Foster & Skehan, 1996; Skehan & Foster, 1997), and some learners (e.g. Kawauchi, 2005; Tavakoli & Skehan, 2005; Wigglesworth, 1997) only. In the studies on planning in L2 written production (e.g. R. Ellis & Yuan, 2004; Ojima, 2006), no positive effect of pre-task planning on accuracy was found. Results of these studies indicate that pre-task planning has limited effect on language production. However, given the fact that there are some studies that have found some effects for pre-task planning on accuracy in oral production and that the number of studies that were conducted in a written context is small, there is not yet enough evidence to conclude that pre-task planning does not influence the accuracy of L2 production. Besides, previous studies have found that some factors, such as allocation of attentional resources, task variables, learner orientation, and problems with measurement, influence the results for accuracy. Some of these are hypothesized to affect the results for accuracy in the present study as well, and will be discussed further below.

#### *Learner orientation*

Learner orientation refers to how learners approach the task. Here it refers to how

participants fulfilled the tasks according to their understandings of task instructions.

For each task in the present study, a list of words and expressions was provided on the task sheet for the purpose of eliciting more comparative forms and suggesting possible ideas for inclusion (see Chapter 4 section 4.5.1). The list was termed on the task sheets ‘useful words and expressions’, suggesting that it was there to provide help. However, some participants understood the instruction as requiring them to include the words in the essays. Four pre-task planners revealed in the post-task interview that they had spent planning time thinking about how to use the words on the list. Though only four students explicitly stated that they made an effort to include the words on the list in their essays, this understanding of task instruction could be inferred in comments of some other students. For example, one student reported in the interview that “I thought the tasks were difficult for me because I did not know many of the words on the list.” It can be inferred from this student’s comment that he took it as a requirement to include the words on the list in the essays and this caused difficulty for him. Since this understanding of task instructions was not what the researcher had expected, there was no question designed in the interview to elicit this information. Therefore, data on this was volunteered by the interviewees and was incomplete. To ascertain that learners’ understanding of task instruction was a factor that influenced the results for accuracy, the number of words on the lists that were used in the essays was investigated. To do this, the words on the lists were compared with the essays using Text\_Lex compare program (see Chapter 4 section 4.7.1.2 for a description of the program). Results show that for the pre-task planning group on average over 50% of participants used more than two thirds of the words on the list and 72% of participants used more than half of the words on the lists. Therefore, it is quite evident that there was a discrepancy between the intention of the task designer (i.e. the researcher) and some students’ understanding of task instructions.

The result of this discrepancy is that the list of words and expressions, on the one hand, might have directed students’ attention to form in terms of lexis and morphological forms (e.g. cheap—cheaper), which could have facilitated use of comparatives. Nevertheless, as will be discussed in more detail in the next section, this aid did not help

increase accurate use of comparative forms significantly since the students could already achieve a high rate of accuracy at the outset of the study, leaving little space for improvement. On the other hand, the unfamiliar words on the list placed extra linguistic demands on the students when they tried to include these words in their essays. This might have increased the task difficulty (i.e. vocabulary load in the code complexity dimension in Skehan's theory for task complexity discussed in Chapter 2) for some learners and caused a decrease in accuracy. One example of an unfamiliar word causing errors was the word 'eye-opening'. Ninety five percent of participants in PTP used it in their essays but 90% of them did not use it grammatically correctly. For example, one student wrote "studying abroad not only make me *eye-opening* but also open-minded". Another wrote "I would become *eye-opening* and open-minded". In short, when students thought they were required to include as many words on the list as possible in their essays avoidance of unfamiliar words would be difficult. The chance to plan seemed to have drawn learners' attention to these words while decreasing the attention they could have otherwise paid to other linguistic forms needed for task completion and the effort of using unfamiliar words in their essays caused errors. These factors affected accuracy of texts. This finding supports Ortega's (Ortega, 1999, 2005) argument that learners' interpretation of task requirements could be a moderating factor for the effects of pre-task planning on task performance. It is also in line with the distinction between *task* and *activity* in sociocultural theory. According to this theory, "the task represents what the researcher (or instructor) would like the learner to do, and activity is what the learner actually does" (Roebuck, 2000, p. 84). When doing tasks, individual learners would act according to their own intentions and interpretation of task requirements. Various factors, such as social context and their past experience, are usually at work in shaping how learners are oriented.

In order to understand why some participants understood help given for a task as a task requirement, some background information is relevant. The participants of the present study had received exam-oriented education in the past 12 years. At the time data was collected for this study, this group of learners had just passed the college entrance

exam. The influence of their training for the exam was evident. In the post-task interview, 50% of the pre-task planners interviewed kept mentioning the exam and talked about how they were trained to write for the exam. One of the things they had learned from exam training was that words and information provided on the writing task sheet needed to be included in the essay. If they missed any point required to be included they would be penalized. It is most likely that this understanding of writing tasks was carried over to their university English study. With this understanding, many of them interpreted the useful words and expressions provided on the task sheets as required vocabulary to be included in the essay. This supports Batstone's (2005) argument that social context and learner factors are important aspects that should be taken into consideration in planning studies.

#### *Limitation of measurement*

In addition to the reasons mentioned above, limitations with the measures used in the present study could also be a possible explanation for the lack of effect on accuracy. Problems exist with both specific measures and general measures for accuracy used in the present study.

For the specific measures, it was hard to decide which linguistic form(s) to zero in on. Previous studies that used specific measures made decisions on which form(s) to examine mainly based on such considerations as whether the form(s) posed difficulty for the learners, whether learners were likely to have acquired the form(s) according to Pienemann's learnability and processability theory (Pienemann, 1989, 1998), and the extent to which a sufficient number of occurrences of the form(s) can be elicited. Following previous studies, the present study selected comparative forms and subject-verb agreement in consideration of the above-mentioned factors. However, the two target forms seem problematic as measures of accuracy because the accuracy levels were already quite high in writing at the outset of the study, contrary to the expectations of the researcher. The mean accuracy scores for the comparatives in the pre-test writing task were: NP: 95%, OLP: 93%, and PTP: 90%. The mean scores for S-V agreement in the pre-test writing task were: NP: 95%; OLP: 96%; and PTP: 96%. Because learners

could already perform at a high level in the use of the target forms at the outset of the study, it left little room for improvement. Improvement in the use of the two forms could hardly achieve statistical significance.

As for the two general measures (i.e. *errors per 100 words* and *ratio of error-free clauses*) employed in the present study, they might not be sufficiently sensitive to detect changes over a short period of time. While widely used, they have been criticized for being too broad to capture small changes in interlanguage (Ortega, 1999), because they, though covering a broad range of error types, obscure errors that feature at a particular developmental stage (Bardovi-Harlig & Bofman, 1989; Ortega, 1999; Wolfe-Quintero et al., 1998). It has been acknowledged that IL development requires long-term restructuring and refining of the internal system. It might be too ambitious to expect a significant increase in measures for overall accuracy after a short planning time. Moreover, the two indices were calculated on the basis of instances of errors and information regarding types of errors (e.g. lexical errors or grammatical errors) and seriousness of errors, referred to as error gravity (C. James, 1977; Vann, Meyer, & Lorenz, 1984), was not available from these general measures. It is possible that pre-task planning may have been able to have an impact on certain types of errors and reduce severity of errors in the present study. Types of errors and seriousness of errors were not assessed in the present study, thus the picture obtained regarding accuracy was not complete.

#### *Operationalization of pre-task planning*

As has been reviewed in Chapter 3, though many previous planning studies did not find an effect for pre-task planning on accuracy, a few studies did find increased accuracy in some measures as a result of pre-task planning (e.g. Foster & Skehan, 1999; Mochizuki & Ortega, 2008; Skehan & Foster, 2005). A common feature of these studies was that a high degree of guidance on what to plan and how to plan was given to learners. The participants in the pre-task planning groups were either given specially designed guidance notes (Skehan & Foster, 2005) or explicit teaching of the target forms prior to planning (Foster & Skehan, 1999; Mochizuki & Ortega, 2008). This indicates that the degree of guidance provided may have an impact on task performance in terms of

accuracy. In the present study, participants in the pre-task planning group were not provided with guidance notes on how to plan in detail. This could be a possible reason for the lack of effect of pre-task planning on accuracy.

In summary, as in many previous studies on pre-task planning, the present study did not find positive effects for pre-task planning on accuracy. However, it is not possible to conclude yet that pre-task planning does not affect accuracy of L2 writing until the possible impact of variables such as learner orientation to task and lack of planning guidance have been investigated. Future research also needs to carefully choose the measures of accuracy.

#### **5.6.1.2 The effects of online planning on accuracy**

In addition to pre-task planning, the present study also investigated the effects of online planning on L2 writing. This was to probe into the different functions of pre-task planning and within-task planning (R. Ellis, 2005b) in the writing process and to test Wendel's hypothesis that online planning might be able to increase accuracy of language (Wendel, 1997). According to him, off-line planning (pre-task planning) could not promote grammatical accuracy for speech production because during the off-line planning period speakers could not "anticipate the item-by-item, moment-to-moment grammatical requirements of the intended communication" (Wendel, 1997, p.143). In contrast, during online planning speakers could both plan and monitor their speech. As a result, online planning, interacting with monitoring, promotes accuracy. This hypothesis has attracted the attention of scholars seeking explanations for the inconsistent results for the effects of planning on accuracy. A few studies have tested this hypothesis in the context of oral production (e.g. Nakakubo, 2011; Wang, 2009; Yuan & Ellis, 2003) and the findings were mixed. Though it has been acknowledged that online processing pressure in writing is much less than in speech, some scholars have investigated the effects of online planning on written production as well, believing that Wendel's hypothesis might also apply in written contexts. R. Ellis and Yuan (2004) were the first to test this hypothesis in a written context and they found that online planning led to a



higher accuracy score in one measure for overall accuracy (i.e. percentage of error-free clauses) and one specific measure for accuracy (i.e. correct verb forms). The results of this study indicate that Wendel's proposal might be relevant for writing as well and encouraged more studies to investigate the role of online planning in written production. Several studies (e.g. Al-Humaidi, 2008; D. Li, 2004; Rahimpour & Safarie, 2011;) have been conducted since then but no significant effect of online planning on accuracy was found. The fact that most of the studies that investigate the effect of online planning on written production to date have not found a significant impact on accuracy (except Ellis & Yuan's study) suggests that online planning might bring limited benefit to accuracy of written production.

The results of the present study are consistent with those of the studies discussed above but run contrary to R. Ellis and Yuan's study (2004). One important difference between the present study and R. Ellis and Yuan's study is the operationalization of the online planning condition. In their study, the online planning group was given unlimited time to perform the task whereas in the present study there was a time limit for the online planning group (25 minutes), though it had more time to write than the other two groups (15 minutes). Since the learners in the online planning group had unlimited time in R. Ellis and Yuan's study (2004), they had more opportunities to engage in both pre-task planning and online planning. It was not clear that the positive effects on accuracy found in their study were caused by online planning or by a pre-task planning +online planning condition. It might be due to this difference in operationalization of online planning that the results of the present study run counter to R. Ellis and Yuan's study (2004).

One possible explanation for the ineffectiveness of online planning in the present study was that lifting the time pressure alone could not lead to accuracy. Though the online planners had more time to write than participants in the other groups, they also had to deal with conceptualization, formulation, and monitoring simultaneously. How to make use of the extra time and how to allocate their attentional resources were completely in the hands of the participants. For example, the interview data revealed that 54% of the online planners reported having paid attention to complexity when writing

while only three students reported having paid attention to accuracy. It seems that providing more time for learners to produce output has to be combined with other intervention measures, such as directing their attention to correctness of linguistic forms, to increase accuracy of output. This view could find some support from Hulstijn and Hulstijn's study (1984), which investigated the influence of time pressure and focus of attention on the correct use of two grammatical forms. They found that attention had a significant effect on grammatical correctness of the target forms but time pressure did not. They concluded that time in itself was not a necessary condition for successful self-correction, which contributed to accuracy, but focusing on language needed time to bring about successful self-correction. Although they did not set out to examine the effect of online planning on accuracy, their findings support the argument that providing more time alone will not bring about benefit to accuracy.

The result of the present study, together with the findings from other studies that did not find a positive effect of online planning on accuracy, poses a challenge to Wendel's hypothesis. It points to a conclusion that providing more opportunities to do online planning alone might not necessarily lead to an increase in accuracy. This is because how to make use of the extra time was up to the learners to decide. If learners decide to devote the time to working on the content of production or focusing on other aspects of language, as in the case of the present study (more than half of the participants chose to focus on complexity rather than accuracy), a gain in accuracy will not be evidenced.

Another factor that might have influenced the results for effects of online planning was that students' existing L2 knowledge put a ceiling on the effects on accuracy. Though participants in the online planning group in the present study could make use of the extra time they had to monitor and edit output, monitoring and editing could only help reduce errors they had correct explicit knowledge of and those caused by carelessness. For errors that could not be identified with their current proficiency or knowledge, monitoring and editing could not help. In the post-study interview, 31% of the online planners commented that they were not able to locate errors in their own writing though they had time to edit. For the rest of the online planners who reported having time to edit their

essays, only one or two errors were corrected and they were all problematic grammatical areas the participants were aware of, such as errors in tense or 3<sup>rd</sup> person 's'. This indicates that for some participants their proficiency level put a ceiling on the number of errors they were able to reduce with the opportunity to monitor and edit. Similar complaints from participants regarding a ceiling for the effect of planning are also evidenced in other studies (e.g. Ortega, 1999). Ortega (1999), when proposing learner proficiency to be one important factor influencing the effects of planning, argued that learners with low degree of control over the target language and with incorrect L2 representations may not be able to benefit from extra time alone without appropriate L2 assistance. The results of the present study, to a certain extent, support this argument.

Apart from the explanations outlined above, learner orientation to tasks and problems with measurement that have been discussed in the previous section might also apply in accounting for the lack of effects for online planning.

## **5.6.2 Complexity**

In this section, the effects of planning on both syntactic complexity and lexical complexity will be discussed.

### **5.6.2.1 Effects of pre-task planning on complexity**

That the effects of pre-task planning did not show in syntactic complexity runs counter to many previous planning studies on oral production (Foster & Skehan, 1996; Guar-Tavares, 2008; Mehnert, 1998; Ortega, 1999; Skehan & Foster, 1997; Wang, 2009; Wendel, 1997; Williams, 1992). It seems that the majority of the previous studies found syntactic benefits for pre-task planning, though a few studies (Mehnert, 1998; Rutherford, 2001; Tajima, 2003) failed to find a significant increase in syntactic complexity in the planning condition.

The difference between the results of the present study and those of the planning studies on oral production could perhaps be explained by the differences between the two modes of production. Although it could be argued that processes involved in speaking

and writing have much in common (R. Ellis & Yuan, 2004; Kellogg, 1996), it is widely acknowledged that writing is less time constrained (Kormos, 2011). When writing, L2 learners can focus on one stage of processing at a time because they are not under real-time communication pressure to simultaneously plan and linguistically encode the message. In contrast, the communication pressure in oral production is greater and learners have to conceptualize, formulate, and articulate simultaneously. Because of those differences it is reasonable to hypothesize that planning might play a different role in writing than in speaking. Pre-task planning enables learners to prepare the propositional content and the language needed for meaning conveyance (at least partially) before they perform the task, and thus reduces the online processing load so that learners could attempt greater challenges during task performance (Foster & Skehan, 1996). This might lead to greater complexity in oral production. However, as has been explained above, the online processing need is less pressing in writing compared to speaking. Therefore, in written production the advantages brought about by having the opportunity to conceptualize and formulate prior to task would not be as strong and evident as in oral production.

While the results of the present study on syntactic complexity were different from most of the previous studies on planning in oral production, they were consistent with most of those on planning in L2 writing. Actually, only a few studies have been found so far to have investigated the effects of pre-task planning on L2 written production. Among the nine studies reviewed in Chapter 3, one study (Ong & Zhang, 2010) did not measure syntactic complexity, thus a comparison of the results of the present study and those of Ong and Zhang's study was not possible. In the other studies, five did not find any gain in syntactic complexity under the pre-task planning condition (e.g. Li, 2004; Johnson, 2011). One (Ojima, 2003) was a case study with only three participants. Though the results of this study showed that planned compositions were syntactically more complex than the unplanned ones, the finding cannot be generalized to a larger population because of the small sample size. Two studies (R. Ellis & Yuan, 2004; Farahani & Meraji, 2011) found increased syntactic complexity under the pre-task planning condition. R. Ellis and Yuan's

(2004) is the one most comparable with the present study because the two studies shared similarities in the independent variables (i.e. planning conditions under investigation) and dependent variables. They concluded that pre-task planning had some effect on syntactic complexity based on the finding that pre-task planners produced greater variety of verb forms, which is a measure for syntactic variety (not measured in the present study), than the non-planners. With regard to the other measure for syntactic complexity, i.e. number of clauses per T-unit, no significant between-group difference was found. Their result for the number of clauses per T-unit fits with the result obtained in the present study.

According to results of the previous planning studies and those of the present study, it could be concluded at this stage that pre-task planning seems to have little effect on syntactic complexity of written production while its effect on syntactic complexity of oral production is stronger.

Nevertheless, a number of factors could have influenced the results for complexity in the present study. The first factor is the task effect. As has been discussed in Chapter 4, the cognitive demands of the writing tasks in the present study were reduced by task supporting elements, such as provision of background information and task structure. It is possible that the tasks in the present study were not complex enough for an effect of pre-task planning to show. This hypothesis is made on the basis of previous studies that have found task features to be influential in task performance (Farahani & Meraji, 2011; Gilabert, 2007; Kormos, 2011; Kuiken & Vedder, 2008; Ong & Zhang, 2010; Robinson, 1996, 2001, 2005, 2007). According to some scholars (Robinson, 2001, 2003, 2005; Robinson & Gilabert, 2007), provision of pre-task planning time can reduce task complexity and ease the cognitive load on learners (Skehan, 1996a, 1998), thus leading to a better performance. However, the effects of pre-task planning would be more evident in cognitively more challenging tasks (e.g. Foster & Skehan, 1996; Kellogg, 1990; Wigglesworth, 1997).

This hypothesis is supported by Kellogg's study (1990), which has been discussed in Chapter 3, on the effectiveness of pre-writing planning on L1 writing quality. In his study, two prewriting strategies were investigated: outlining and clustering. Another variable

that was examined was the demands of the writing tasks controlled by three task conditions: topic only condition, topic + ideas condition, and topic + ideas + organization. In the topic only condition, the task demanded generation and organization of ideas. In the topic + ideas condition, the task only demanded organization of ideas. In the topic plus ideas plus organization condition, a suggested organizational scheme in addition to the topic and the relevant ideas was provided. The cognitive demand in the topic only condition was supposed to be the highest and lowest in the topic plus ideas plus organization condition. The study found that outlining significantly improved the overall quality of essays measured in terms of content (i.e. how well the ideas were developed, how coherent the text was, and how effectively the message was conveyed) and style (i.e. quality of word choice, sentence structure, spelling, and grammar). The beneficial effect was the strongest in the topic only condition. As the task demands reduced, the effect of outlining weakened and the effect disappeared in the topic plus ideas plus organization condition. Though Kellogg's study was conducted in an L1 writing context, which is different from an L2 context, it has important implications for studies carried out in L2 writing contexts. This is because there are some similarities between L1 and L2 writing strategies (Sasaki, 2000). The writing tasks in the present study, which only required putting together information as understood by some participants, were similar to the topic plus ideas plus organization condition in Kellogg's study. Therefore, they may not be complex enough for the benefits of pre-task planning to show, especially in terms of accuracy and complexity.

More specifically, the provision of information on the task sheet may have influenced the results for syntactic complexity. As has been described in Chapter 4, for each task some information on the objects under comparison was provided to alleviate the difficulty of idea generation under time pressure. Some participants basically followed the information provided on the task sheets without bothering to be more creative in term of content. To find out the extent to which participants copied the information provided, a comparison of the information on the task sheets and the content of the essays was conducted. Every point made in an essay was looked at to see if the information was from

the task sheet. From the comparison, it was found that on average 76.7% of the points made in the essays (the lowest percentage being 67.5% and the highest being 87%) were from the given information on the task sheets. Thus, the essays produced by participants were similar in terms of content and this might have led to similarities in syntactic complexity for the three groups. This explanation could find theoretical support from the studies on task complexity. Based on the assumption that complex concepts require the use of complex syntactic structures (e.g. Robinson's Cognition Hypothesis, 2001, 2003, 2005), Kormos (2011) hypothesizes that task complexity of L2 writing tasks makes demands on the planning of the content of the text and the content in turn determines the linguistic encoding that takes place at the formulation stage. Because of the provision of information on the task sheets, the demand for planning the content was reduced and the pre-task planners did not have to spend much time generating ideas. The fact that most of the ideas were copied from the provided information made the content of the essays produced by the three groups very much similar. The advantage of pre-task planning in terms of conceptualization was weakened.

The second factor influencing the results could again be the limitations with measurement. The present study used two widely employed measures for syntactic complexity: number of clauses per T-unit (C/T) and mean length of T-units (MLT). Though they are popular in studies on L1 and L2 writing, the findings so far cannot lead to the conclusion that they can well indicate syntactic development of learners over a relatively short time. When discussing these popular measures for syntactic complexity, Ortega (2003) pointed out that an observation period of roughly a year is needed for syntactic development of college-level learners to be observed, suggesting that these measures for syntactic complexity might not be sensitive enough to changes over a short time. Moreover, Biber et al. (2011) argued, on the basis of a corpus-based analysis of professional academic papers, that measures for T-units and clausal subordination are more characteristic of conversation than academic writing. Their finding could partly explain why the effects of pre-task planning were more evident in oral production than written production in terms of syntactic complexity. Some other studies also provided

evidence to show that number of clauses per T-unit might not be sufficiently sensitive to detect differences in syntactic complexity in written production. Knoch's study (2007), for example, used number of clauses per T-unit to assess writing scripts at different DELNA (Diagnostic English Language Needs Assessment) writing levels and found that it failed to differentiate scripts at different ability levels. Lu (2011) also found, through a corpus-based evaluation of 14 syntactic complexity measures used to assess college-level English writing of Chinese learners, that clauses per T-unit was not able to discriminate essays written by students of different proficiency levels determined by programs. It seems that clauses per T-unit is problematic as a measure for syntactic complexity of written scripts.

With regard to the other measure used in the present study, mean length of T-unit (MLT), it has been used widely in L1 and L2 writing research to measure overall syntactic complexity since it was developed by Hunt in 1965. It has been proved as generally a valid measure for L1 language development, in which the development of the ability to use more complex stylistic devices was in a linear manner (Gaies, 1980). In second language acquisition, however, a non-linear development in complexification has been argued (Cooper, 1976, Monroe, 1975, Wolfe-Quintero et al. 1998, cited in Ortega, 2003). For this reason, such a measure of syntactic complexity might not be able to differentiate performances under different task conditions. In addition, T-unit analysis could not capture complexity at phrasal level, as suggested by Lu (2011). To capture phrasal elaboration, Norris and Ortega (2009) proposed that mean length of clause be used to complement measures for global syntactic complexity (e.g. MLT) and complexity by subordination (e.g. C/T). In this way, a better picture of how L2 syntactic complexity develops could be seen.

In terms of lexical complexity, two measures were used in the present study: lexical variety (D) and lexical sophistication (Lam). No significant between-group difference was found in D but a significant difference between PTP and NP was detected in Lam ( $p=.023$ ). Lexical sophistication is a measure not frequently used in previous studies and a significant advantage of the pre-task planning condition was found. This indicates that



pre-task planning offers more opportunities to retrieve lower frequency words in written production. According to VanPatten (1996), learners pay attention to lexical items before they attend to grammar when processing meaning. It was not surprising that, with the extra time to plan, the pre-task planners focused on retrieving lexis that could better express their ideas, which resulted in producing lexically more sophisticated essays than the no planning group. Evidence from interview data in the present study suggests that participants regarded using more advanced words as one way to impress the reader. For example, one student reported, “I spent time on thinking about more advanced words to use because this will make the essay good.”

With regard to the results concerning lexical variety (D), they were consistent with two of the three studies that used this measure to investigate the effects of planning on L2 writing (R. Ellis & Yuan, 2004; Johnson, 2011) but different from the other study (Ong & Zhang, 2010). R. Ellis and Yuan (2004) and Johnson (2011)’s studies did not find significant between-group differences in lexical variety whereas Ong and Zhang’s study (2010) found that the free-writing and no planning condition achieved significantly higher scores for lexical variety than the pre-planning condition. In a sense, the results for lexical variety of the present study and the three studies mentioned above did not differ in that none of the studies found positive effects of pre-task planning on lexical variety.

There is again an issue of measurement. Skehan used both lexical variety (D) and lexical sophistication (Lambda) in his study (Skehan, 2009a) to measure the effect of planning on lexical performance of both native and non-native speakers of English in speaking. He found that the difference between native and non-native speakers was more evident in Lambda than in D. The correlation between the two measures was very low, which was against the possible speculation that D would impact on Lambda. In terms of the effects of planning, the results were consistent across L1 and L2 speaking. There were no significant results regarding D but the results for Lambda were significant. The results of the present study in the two measures fit with his study (Skehan, 2009a). It should be noted that in most previous planning studies significant gains in D were not detected either. This leads to a conclusion that lexical benefit from planning is more likely to be

detected by Lambda than by D. This could again be discussed in relation to the task characteristics. As described earlier, participants basically used information provided on the task sheets to fulfill the tasks. That is, the content of essays was prescribed and participants were pushed to retrieve less frequent lexis required by the tasks. In this case, precision might be prioritized over avoidance of recycling words (D measures the extent the writer recycles words in the essay). The question of what measures are best for measuring planning influenced lexical performance needs to be answered by future studies.

There is another interesting finding in Skehan's study (2009a). He found that Lambda correlated negatively with syntactic complexity, which means that the use of less basic words led to lower syntactic complexity. He hypothesized that less common words posed difficulty on L2 users at the formulating stage of speech, which disrupted syntactic planning. Therefore, there was a toll (i.e. less complex sentences) for those who mobilized more challenging words in production. If he is right, then the higher rate of using more advanced words by PTP could be another reason for the lack of gain in syntactic complexity.

To conclude, no evidence has been found in the present study to show a positive effect of pre-task planning on syntactic complexity. Task features and problems with measurement have been found to be the main influential factors. In terms of lexical performance, a positive effect was found for lexical sophistication by not for lexical variety. An issue of measurement was forwarded to explain this result.

#### **5.6.2.2 Effects of online planning on complexity**

In the present study, online planning did not bring any benefit to syntactic complexity either. Though participants in the online planning group had more time to generate ideas than those in the no planning group, they might not have done so. As has been explained in the previous section, the provision of information on the task sheets reduced the necessity of idea generation and most of the participants just made use of the information to complete the tasks. This caused similarity in content of the essays

produced by all three groups. Language complexity is to a large degree determined by complexity in content (Robinson, 2001, 2003, 2005; Kormos, 2011), therefore, the three groups produced almost identical scores in syntactic complexity.

Another explanation could be that the extra time online planners had to write could not help lift the ceiling placed by their limited L2 knowledge. According to the Output Hypothesis (Swain, 1985, 1995, 2000; Swain & Lapkin, 1995), producing output can help learners notice a gap between what they need or want to say and what they can say. However, having extra time to produce language could not help fill the gap they have noticed. In the post-task interview (see Chapter 7 for results), one of the online planners commented: “I wanted to write some complex sentences and I tried. But the number of sentence patterns I have at disposal is limited”. Another online planner reported, “sometimes I tried to use a some complex sentence but then I found that I did not have the command of it so I had to give up.” Two other online planners complained that they felt frustrated for not being able to use more variety of language: “I felt fed up with my using vocabulary and sentence patterns repetitively but I could not come up with other ways to express the ideas”. Because having more time to produce language alone could not help fill the gap between what they wanted to say and what they could say, to fulfill the experimental tasks of this study, many participants just expressed the content prescribed by the task in whatever L2 linguistic resources they had available. This might have put a ceiling on what the extra time could buy.

The limitations with measurement already discussed in the previous section also apply in accounting for the lack of effect of online planning on syntactic complexity.

As with previous studies on planning effects on L2 writing performance, no significant effect was found in lexical complexity for online planning in the present study. While some pre-task planners reported having made efforts to use more advanced words, no such report was found in online planners. Thus, online planning did not bring any benefit to lexical sophistication. As explained in 5.6.1.2, the online planners had to deal with conceptualization, formulation, and monitoring simultaneously while writing. They might not have extra attentional resources for retrieving more advanced lexis, which

might be the cause for the lack of effect on lexical sophistication. The possible reasons proposed for lack of advantage for pre-task planning on lexical variety might also account for the nil effect of online planning.

### **5.6.3 Fluency**

#### **5.6.3.1 Effects of pre-task planning on fluency**

That pre-task planning did not significantly impact temporal fluency was inconsistent with findings in previous studies on planning in oral production (e.g. Foster & Skehan, 1996, 1999; Gilabert, 2007; Mehnert, 1998; Ortega, 1995b; Tajima, 2003; Wendel, 1997). The difference between the results of the present study and those of the studies on planning in oral production could be explained by the differences in modality. In speaking, the speakers are under great pressure to simultaneously engage in conceptualizing, formulating and articulating. Thus, having the opportunity to conceptualize and linguistically encode the ideas during pre-task planning time can evidently aid fluency of speech. In writing, there is less real time communicative pressure and the writer could focus on one stage of processing at a time. Therefore, the advantage of pre-task planning on writing brought about by prior conceptualization and formulation might not be as evident as on speaking.

In addition to the differences in processing between speaking and writing, the length of task completion time might also have made the results found in studies on oral production and those of the present study different. An oral task, particularly a monologic task, usually only takes a few minutes to complete while a written task would usually take as long as 20---60 minutes to finish. In Skehan and Foster's study (2005), it was found that the effects of pre-task planning significantly reduced after 5 minutes. It seems that learners can only hold in their memories their plans for a short time and have to depend on improvisation after the pre-task plan fades. Though this was the only study that examined the durability of planning effects (i.e. how long the impact of pre-task planning could be sustained), it offers insights into this important issue. It also provides

an explanation for the differences found in studies on planning in writing and those on planning in speaking.

Comparing with other studies on planning in written production, it was found that the results of the present study regarding temporal fluency were different from six studies (Al-Humaidi, 2008; R. Ellis & Yuan, 2004; Farahani & Meraji, 2011; M. D. Johnson, 2011; Ojima, 2006; Rahimpour & Safarie, 2011) and consistent with three other studies (D. Li, 2004; Ong & Zhang, 2010; Pu, 2009). The difference between the results of the present study and these six studies might be caused by task type. Most of these studies (e.g. Al-Humaidi, 2008; R. Ellis & Yuan, 2004; Farahani & Meraji, 2011) used narrative writing tasks while persuasive tasks were used in the present study. Ong and Zhang's study (2010) used argumentative writing tasks, which were similar to the tasks used in the present study, producing similar results for temporal fluency as those of the present study. This indicates that task type might have led to the difference in the results of the present study and other studies that found increased fluency in writing. It is possible that the pre-task plan of a narrative discourse is easier to be recalled in subsequent writing than that of an argumentative discourse.

Studies on memory found that the gist of a text is more easily memorized than the precise wording (Brainerd & Reyna, 1998; Roediger & McDermott, 2000). The precise wording is important for argumentative writing (Britt, Kurby, Dandotkar, & Wolfe, 2007) because a subtle change in wording will significantly change the argument. For example "death penalty should be abolished" would be different from "death penalty is abolished". Britt et al.'s study (2007) also showed that predicates of arguments were more difficult to recall than those of narrative statements. Given the importance of careful wording in this type of writing and the difficulty of verbatim memory, it is speculated that the pre-task plan of an argumentative discourse is more difficult to recall and execute during task completion than that of a narrative discourse. This difficulty in recalling and executing the pre-task plan might influence the speed of writing. In the case of the present study, though the participants could use the information provided on the task sheets as prompts for what to write, they still had to rely greatly on their memories for the precise wording

and predicates to use. When they tried to recall the precise wording, their writing speed might be affected. What is presented above is just a hypothesis. Why pre-task planning should have more effects on fluency in narrative writing than on persuasive/argumentative writing tasks needs further research to explain.

That pre-task planning did not help pre-task planners write significantly faster than students in the control group could be a result of the task features. As has been described earlier, there was some information provided on the task sheets that could be used when writing. This seemed to have reduced the potential advantage of pre-task planning on conceptualization. Theoretically, pre-task planning provides opportunities for learners to plan and organize the propositional content prior to the task, thus reducing the online processing load. As a result, the speed of production is increased because with a planned content and structure learners should be able to speak or write faster. Moreover, it is reasonable to expect this possible effect of pre-task planning to be more evident in tasks with high conceptual demands. In the case of the present study, the task demand on conceptualization was reduced by the provision of information that participants could make use of in their essays. That is, with the information provided on the task sheets, all participants did not have to spend much time and effort conceptualizing the content. Therefore, the potential effect of pre-task planning on temporal fluency was not evident.

While the impact of pre-task planning on temporal fluency was not significant, its impact on dysfluency was significant. In this regard, the result of the present study was consistent with R. Ellis and Yuan's study (2004). It also corroborates Kellogg's study (1987) on planning in L1 writing, which showed that subjects in the outline condition did less reviewing than those in the no outline condition. This could be explained by the fact that pre-task planning allowed the learners the opportunity to plan for both content and language. According to their planning notes and their answers to the post-task questionnaire and interview, the majority of the pre-task planners engaged in content and organizational planning as well as language planning during planning time (see Chapter 7 Section 7.3 for details). This would result in alleviation of on-line processing load (R. Ellis & Yuan, 2004) and gave them confidence in writing. Moreover, language planning

gave learners an opportunity to rehearse the language they were to use in their writing. During production, they could use what they had rehearsed and thus the pressure of monitoring was lessened. Zimmerman (2000) found that more revisions were made in L2 writing than in L1 writing. Pre-task planning seemed to have the benefit of compensating for the lack of proficiency in L2 writing and resulted in more L1 like writing behavior in the case of repair fluency. This is one possible reason why the pre-task planning group was more fluent in writing in terms of dysfluency.

### **5.6.3.2 Effects of online planning on fluency**

Similar to R. Ellis and Yuan's study (2004), the present study found that online planning caused a decrease in writing speed and number of self-corrections. This result is also consistent with Yuan and R. Ellis (2003) and Wang's (2009) studies in that Yuan and R. Ellis (2003) found that the online planning group had a significantly lower score for speech rate than the pre-task planning group and Wang (2009) found the watched and online planning group (the watched condition was considered a pre-task planning condition, in which the participants watched the video once before they narrated the story) had a large effect on repair fluency.

The decreased writing speed could be attributed to the fact that the online planning group had more time to complete the tasks than the other groups. Having extra time to write means two things: less time pressure and more opportunities to monitor their output during writing. Scores given by the three groups in the post-task questionnaire regarding whether or not they felt pressed for time were: 3.03 for PTP, 2.7 for OLP, and 3.47 for NP. The higher the score the more pressed for time they felt. Thus, it could be seen that among the three groups OLP felt the least pressed for time. According to Nakakubo's study (2011), time pressure is a factor that significantly affects fluency of production. The speed of production of the participants in the time-pressured groups was significantly faster than that of the participants in the non-time-pressured groups in her study. This finding was not surprising because when learners try to complete a task within a time limit they tend to speak or write fast. Though Nakakubo's study was conducted in a

speaking context, it might be applicable in a writing context. Therefore, it could be argued that one possible reason for the low writing speed of the OLP was that they had less time pressure than the other two groups.

The second consequence of having more time to write was that participants in the OLP were more likely to monitor their output. According to the results of the post-task interview, more participants in the OLP reported having paused, reread, and edited during writing than the other two groups (see Chapter 7 Section 7.4.3 for details). This indicates that more students in this group engaged in online monitoring. For writers who mainly rely on controlled processing, as is the case with participants of this study, monitoring might disturb writing speed. As Kellogg (1996) explained, pauses and monitoring during formulation should greatly slow down the speed of text production.

Overall, pre-task planning aided fluency in writing and online planning had negative effects on writing speed.

## **5.6.4 Organization**

### **5.6.4.1 Effects of pre-task planning on organization**

Analysis of discourse features of the essays measured by subjective ratings and computer program generated scores for cohesion and coherence showed that there were no significant differences among the three groups. This result is consistent with Kellogg's study (1987) on the effects of pre-writing strategies on L1 writing, in which he found that letters produced in the outline conditions were not significantly different from those produced in the no outline conditions in terms of coherence. This indicates that producing a cohesive and coherent text might be related to one's discourse competence, which would not change according to task completion conditions.

The present study is not directly comparable to other L2 studies since few previous studies on planning measured discourse features of learners' production. The few studies that did take discourse quality of production into consideration used different measures. Williams (1992) used frequency of discourse markers as a metric to measure discourse



and found planning increased discourse marking. Crookes (1989) also used this metric and found greater use of discourse markers in one of the tasks. Discourse markers are superficial discourse features that are not necessarily significantly correlated with quality of texts (Liang, 2006; Ma, 2002). Some L2 learners might overuse or misuse connectives in their essays, rendering frequency of occurrence of connectives an inappropriate measure for discourse quality of texts. Therefore, the present study did not use it as an index for organization.

Another relevant study is Wigglesworth's study (1997), which used test scores assigned by two raters to look into the effects of planning on discourse. No significant differences between the planned and the unplanned production were evidenced. The result was attributed to the failure of the raters in perceiving the differences at micro level. The result of the present study was consistent with her study. However, the explanation offered by Wigglesworth was not upheld because cohesion and coherence were measured at both macro level and micro level in the present study and no significant differences were found.

A possible reason for this result could be the design of the tasks. As has been described earlier, some information on the objects for comparison was provided on the task sheets, giving participants the opportunity to simply follow the sequence of the given information when structuring their essays. A comparison of the information on the task sheets and the points made in the essays found that on average 72.6% points presented in the essays followed the order of information provided on task sheets. This might have affected the global organization of the essays in particular.

The second reason could be attributed to the fact that the participants of the present study had limited L2 discourse knowledge because they had not received any formal instructions on how to write essays in English. When giving explanations for their planning and writing behaviors, seven out of fourteen interviewees volunteered the information on how they had been trained to write in English. They commented in the post-task interview that all the writing practice they had in high school was about how to get high scores in the exam. Because in most cases students would be asked to write a

short picture-cued narrative or a few sentences to express their opinions on a certain issue in the exam, they were never taught how to write compare/contrast essays or produce discourse where they had to argue/justify a point in high school. Given this background, it could be said that they had limited discourse knowledge of English essays. Therefore, they were not able to overcome the difficulty of making the text cohesive and coherent even when they had planning time prior to writing.

#### **5.6.4.2 Effects of online planning on organization**

Online planning did not bring any benefit to organization of the essays. One possible reason might be that online planners had to attend to planning, translating, and reviewing (Hayes & Flower, 1980) simultaneously during writing. As Kellogg (1988) argued, only when writers are able to focus on one single process can performance be significantly enhanced.

Apart from the reason mentioned above, the influence of task features and students' approach to tasks together with their limited discourse knowledge (as with explanations forwarded for pre-task planning) might also account for the lack of effect of online planning on organization.

#### **5.6.6 Summary**

Research question one asks whether task planning has immediate effects on L2 written production. Based on the results of the essays written in the experimental sessions the following conclusions can be drawn.

1. Neither pre-task planning nor online planning had immediate effects on accuracy. Several factors have been suggested to influence the results for accuracy. First, learners' understanding of the word list provided for each task might have affected the results. There is evidence to show that some participants regarded the word list as the list of must-use words. This understanding of task instruction made avoidance of using these words difficult and the attention paid to the unfamiliar words on the list might have reduced the attention that could have been paid to other linguistic forms. Moreover, the

unfamiliar words on the lists were sources of errors. Planning opportunities, pre-task planning or online planning, could not help overcome the language difficulties they encountered during production. When students ventured to use the unfamiliar words, errors occurred. Second, there are limitations with measurement. The specific measures for accuracy might not be the best to detect between group differences in accuracy. Also, the measures for accuracy used in the present study were limited to an examination of instances of errors. Types of errors and seriousness of errors were not assessed. Thus, the picture regarding accuracy was not complete. Third, there might not be enough guidance to direct learners' attention to accuracy. For the pre-task planning group, participants were briefly instructed to plan both the content and language. No specific suggestion was given on how to plan the language. For the online planning group, the participants were simply given more time to write. No suggestion was given on how to make use of the time. How to use the time and what to attend to were completely in the hands of the students. Without directing students' attention to correctness of language, planning might not achieve any effect on accuracy. Lastly, it has been suggested that students' limitation of L2 knowledge might have put a ceiling on the effects of planning. For the online planners, even with more opportunities to monitor and edit their output, 31% reported not being able to identify any errors in their own writing. For the rest of the online planners who managed to correct some errors, the number of errors corrected was limited. This suggests that a ceiling is imposed by limited L2 knowledge on the effects of planning on accuracy.

2. Neither pre-task planning nor online planning had immediate effects on syntactic complexity but pre-task planning promoted lexical sophistication. Factors that have been found to influence the results for complexity are: task features, problems with measurement, and limitation of L2 knowledge. There is evidence to show that participants of the present study included a large proportion of the information on the task sheets in their essays. This caused similarity in terms of content, which led to similarity in syntactic complexity. With respect to measurements for syntactic complexity, some studies (Biber et al., 2011; Knoch, 2007; Lu, 2011) found that the measures for syntactic

complexity adopted in the present study might not be good indices that could well indicate syntactic complexity of written texts. There is evidence from the post-task interview to indicate that limitation in L2 knowledge might have influenced complexity of the online planners. Some of them complained that their existing L2 knowledge limit the variety of language they would like to use.

That pre-task planning promoted lexical sophistication suggests that VanPatten (1996) could be right in proposing that learners process lexis prior to grammar when processing meaning. Pre-task planners made use of the planning time to retrieve lexically more advanced words, which resulted in increased lexical sophistication. This also indicates that lexical sophistication could be a good measure that can detect the lexical benefits of task planning (Wang, 2009).

3. Pre-task planning did not promote temporal fluency but enhanced repair fluency. Online planning had negative effects on temporal fluency and had no effect on repair fluency. Task features have been argued to influence the results for temporal fluency of the pre-task planning group. The provision of information on the task sheets reduced the demand on conceptualization and thus reduced the potential effect of pre-task planning on temporal fluency. In the case of online planners, the decreased time pressure and increased opportunities to do online monitoring affected their writing speed. With regard to repair fluency, the chance to rehearse both content and language during pre-task planning alleviated the online processing load and instilled confidence among the pre-task planners when writing. This might have improved significantly this group's repair fluency.

4. Neither pre-task planning nor online planning had immediate effects on organization of the texts. The lack of effect on organization has been explained by the task features again and students' lack of discourse knowledge. Some participants followed the sequence of the given information when organizing their essays. There is evidence to show that many of the points in the essays followed the order of information provided on the task sheets. This might have caused similarity in terms of organization of essays. There is also evidence from the post-task interview that participants had not been

formally instructed how to write essays in English. This indicates that they had limited discourse knowledge, which was a difficulty that could not be overcome by planning opportunities.

## **CHAPTER 6 THE EFFECTS OF TASK PLANNING ON L2 WRITING DEVELOPMENT**

### **6.1 Overview**

The research question this chapter addresses is: RQ2. Does repeated practice under planning conditions lead to changes in L2 writing over time?

To answer this question, a pre-test—posttest---delayed posttest design was adopted. In these tests, all participants were allowed 25 minutes to complete written tasks without pre-task planning time. These tasks were similar to the tasks used in the treatment sessions in design but different in topics (see Chapter 4 Section 4.6.1). The pre-test and the posttest, which was administered in the week following the last treatment session, were four weeks apart, while the posttest and the delayed posttest were ten weeks apart. Participants' performances in the posttest were compared with those in the pre-test to see the effects of planning over a relatively short term while their performances in the delayed posttest (14 weeks after the pre-test) showed the effects of planning over a longer term. As has been explained in Chapter 4, participants' essays written in these tests were analyzed in terms of accuracy, complexity, fluency, and organization. For data that met the assumption of normal distribution repeated ANOVA tests were applied. For data that were non-normally distributed a series of non-parametric tests were undertaken.

This chapter will begin with a report of the results of the pre-test, which showed the participants' writing performance at the outset of the study. This will be followed by the within-group difference of each group in accuracy, complexity, fluency, and organization (Section 6.3). In section 6.4, the performance of the three groups in the testing sessions will be compared to establish whether there were any between-group differences. A discussion of these results will be presented in the last section of this chapter.

### **6.2 Results of the pre-test**

The essays written in the pre-test were analyzed in terms of accuracy, complexity,

fluency, and organization. The descriptive statistics of these measures will be given in the following sections along with descriptive data of the two posttests for convenience of comparison.

To determine if there were any significant between-group differences, statistical tests were conducted.

Table 29 ANOVA results of the general measures in the pre-test

Measures	F	Sig.	Pairwise comparisons: p value (effect sizes)		
			PTP-NP (d)	OLP-NP (d)	PTP-OLP (d)
Accuracy					
error/100	.294	.746	----- (.22)	----- (.00)	----- (.20)
%error-free	.975	.382	----- (.33)	----- (.00)	----- (.31)
Syntactic Complexity					
MLT	1.043	.358	----- (.44)	----- (.29)	----- (.05)
C/T	.721	.490	----- (.35)	----- (.18)	----- (.13)
Lexis					
D	3.565*	<b>.033</b>	.136 (.54)	.807 (.17)	<b>.033 (.72)</b>
Lam	1.03	.361	----- (.43)	----- (.20)	----- (.17)
Fluency					
WMP	2.472	.092	----- (.62)	----- (.46)	----- (.05)
Dysflency	.476	.623	----- (.25)	----- (.21)	----- (.05)
Organization					
Organization rating	5.04*	<b>.009</b>	<b>.016 (.80)</b>	1.00 (.08)	<b>.033 (.72)</b>
Coh_Local	.278	.758	----- (.19)	----- (.18)	----- (.01)
Coh_Global	1.522	.225	----- (.01)	----- (.42)	----- (.45)

Note: Dashes mean that the calculation was not conducted.

Results showed no significant differences among the three groups in the use of comparative forms ( $H(2)=2.594$ ,  $p=.237$ ) or in the use of S-V agreement ( $H(2)=1.058$ ,  $p=.589$ ) in the pre-test. With regard to the general measures of accuracy, complexity, fluency, and measures for organization (Table 29), there were significant between-group differences in two of the 13 measures (i.e. *D* and *Organization rating*) in the pre-test. This indicates that other than lexical variety and organization, participants wrote with a similar level of accuracy, complexity, fluency, and organization, thus leading to the conclusion that their performance in the subsequent treatment sessions and the post-test and the delayed post-test would be comparable.

The fact that there were significant between group differences in *D* and *Organization*

rating in the pre-test scores of these two measures will be taken into consideration in later analysis (see Section 6.4 for details).

### 6.3 Changes over time

In the following sections, the results for accuracy, complexity, lexis, fluency, and organization measures will be reported one after another. Each of these sections will be organized as follows: 1) descriptive statistics of the measures; 2) inferential statistics and effect sizes, and 3) results of post hoc comparisons when a significant difference was indicated in preliminary tests.

#### 6.3.1 Changes in accuracy over time

Table 30 displays the descriptive statistics of the accuracy measures in the three tests. Information on obligatory occasions for suppliance of the target structures is also included in the table for reference.

Table 30 Descriptive statistics for the accuracy measures in the three tests

Measures	PTP (n=25)			OLP (n=25)			NP (n=25)		
	Oblig (M)	M	SD	Oblig (M)	M	SD	Oblig (M)	M	SD
Accuracy									
Comparative									
Pre-test	5.04	.90	.16	4.12	.93	.12	4.24	.95	.12
Posttest1	5.24	.96	.10	5.04	.91	.15	5.12	.92	.14
Posttest2	5.40	.95	.09	4.44	.92	.14	4.88	.95	.13
S-V									
Pre-test	19.36	.96	.06	20.16	.96	.06	18.68	.95	.04
Posttest1	16.64	.96	.04	17.96	.97	.05	18.16	.95	.07
Posttest2	14.88	.94	.09	18.48	.99	.03	16.04	.95	.06
%error-free									
Pre-test		.62	.12		.66	.14		.66	.12
Posttest1		.68	.12		.69	.13		.69	.12
Posttest2		.70	.14		.67	.12		.70	.17
Error/100 words									
Pre-test		.25	.04		.24	.06		.24	.05
Posttest 1		.23	.05		.23	.05		.23	.04
Posttest2		.22	.06		.23	.05		.22	.06

Note: Oblig=mean of obligatory occasions for suppliance of the form.



Table 31 Results for accuracy measures across tests

Measures	Group	Chi-square	Asymp. Sig.	Location of significance: p value (effect sizes)			
				Pre-Post1(r)	Post1-Post2 (r)	Pre-Post2 (r)	
Comparatives	PTP	.844	.656	---- (.23)	---- (.04)	---- (.22)	
	OLP	.174	.917	---- (.00)	---- (.04)	---- (.08)	
	NP	2.52	.284	---- (.24)	---- (.17)	---- (.00)	
S-V	PTP	.078	.962	---- (.03)	---- (.12)	---- (.15)	
	OLP	5.322*	<b>.045</b>	.205 (.25)	.181 (.27)	<b>.016</b> (.46)	
	NP	.000	1.00	---- (.07)	---- (.07)	---- (.07)	
%error-free	PTP	5.515*	<b>.050</b>	.093 (.34)	.396 (.17)	<b>.016</b> (.45)	
	OLP	.388	.824	---- (.19)	---- (.06)	---- (.06)	
	NP	4.061	.131	---- (.11)	---- (.11)	---- (.32)	
Error/100		F	Sig.	Pre-Post1(d)	Post1-Post2 (d)	Pre-Post2 (d)	
		PTP	1.23	.302	---- (.40)	---- (.17)	---- (.50)
		OLP	.243	.785	---- (.20)	---- (.00)	---- (.20)
		NP	.374	.690	---- (.25)	---- (.17)	---- (.33)

Note: Dashes mean that the calculation was not conducted.

Data for the two target structures and percentage of error-free clauses did not meet the assumption of normal distribution, so a series of Friedman tests were run to find out if each group performed at significantly different levels across time (Table 31). Results show that in the use of comparative forms there was no significant within-group difference for NP ( $\chi^2(2)=2.52$ ,  $p=.284$ ), PTP ( $\chi^2(2)=.174$ ,  $p=.917$ ) or for OLP ( $\chi^2(2)=.844$ ,  $p=.656$ ).

In the use of S-V agreement, the changes of NP ( $\chi^2(2)=.000$ ,  $p=1.00$ ) and PTP ( $\chi^2(2)=.078$ ,  $p=.962$ ) over time were not significant. The change of OLP was significant ( $\chi^2(2)=5.322$ ,  $p=.045$ ). The results of the Wilcoxon tests showed that OLP did not change significantly over a period of four weeks (interval between pre-test and the post-test) ( $z=-1.268$ ,  $p=.205$ ). Its performances in the delayed posttest (post2) and the posttest (post1) were not significantly different either ( $z=-1.338$ ,  $p=.181$ ). However, its score in the delayed posttest was significantly higher than that of the pre-test ( $z=-2.318$ ,  $p=.016$ ) with a medium effect size ( $r=.46$ ).

With regard to the percentage of error-free clauses (%error-free), a significant within-group difference ( $\chi^2(2)=5.515$ ,  $p=.05$ ) was found for the pre-task planning (PTP)

group. To find out where the significant difference lay, three Wilcoxon signed ranks tests were run and the significance value was adjusted by dividing .05 by 3 (.05/3=.0167). Wilcoxon test results revealed that the PTP group improved significantly from the pre-test to the delayed posttest (p=.016) with a medium effect size (r=.45) but the improvements between the pre-test and posttest1 and between posttest1 and posttest2 were not significant. The OLP group and the NP group did not improve significantly over time in this measure.

Data for errors per 100 words (error/100) showed that the within-group differences for this measure were not significant (F(2,144)=1.65, p=.196).

In summary, the OLP group improved significantly from the pre-test to posttest2 (p=.016, r=.46) in the correct use of S-V agreement and the PTP group improved significantly from the pre-test to posttest2 (p=.016, r=.45) in the scores for percentage of error-free clauses. The NP group did not improve significantly in any of the accuracy measures.

### 6.3.2 Changes in syntactic complexity over time

Syntactic complexity was measured by mean length of T-units (MLT) and number of clauses per T-unit (C/T). Table 32 shows the descriptive statistics of the two measures.

Table 32 Descriptive statistics for syntactic complexity in the three tests

Measures	PTP (N=25)		OLP (N=25)		NP (N=25)	
	M	SD	M	SD	M	SD
Syntactic complexity						
MLT						
Pre-test	12.87	2.02	12.73	3.34	11.86	2.54
Posttest 1	13.04	2.26	12.56	2.72	12.99	3.63
Posttest 2	12.27	2.90	11.89	2.05	12.36	2.58
C/T						
Pre-test	1.83	.25	1.79	.36	1.73	.32
Posttest 1	1.91	.34	1.89	.29	1.98	.43
Posttest 2	1.80	.33	1.85	.29	1.91	.32

Table 33 Results for syntactic complexity measures across tests

Measure	Group	F	Sig.	Location of significance:		
				<i>p</i> value (effect sizes)		
				Pre-Post1 (d)	Post1-Post2 (d)	Pre-Post2 (d)
C/T	PTP	.807	.452	----- (.08)	----- (.27)	----- (.21)
	OLP	.941	.397	----- (.06)	----- (.33)	----- (.41)
	NP	4.546*	<b>.016</b>	<b>.011</b> (.66)	.496 (.18)	<b>.008</b> (.56)
MLT		Chi-square	Sig.	Pre-Post1 (r)	Post1-Post2 (r)	Pre-Post2 (r)
	PTP	3.84	.15	----- (.07)	----- (.31)	----- (.35)
	OLP	1.04	.595	----- (.04)	----- (.26)	----- (.25)
	NP	3.12	.21	----- (.28)	----- (.21)	----- (.27)

Note: Dashes mean that the calculation was not conducted.

Scores for C/T satisfied the assumption of normal distribution and were analyzed with repeated measures ANOVA but scores for MLT failed to satisfy the assumption and were analyzed with non-parametric tests.

Repeated measures ANOVA results showed that there was a significant main time effect ( $F(2,144)= 4.537, p=.012$ ) for the number of clauses per T-unit. Pairwise comparisons revealed that the learners had a greater number of clauses per T-unit (C/T) in the posttest than in the pre-test ( $p=.005$ ). All three groups produced a higher number of clauses per T-unit in the posttest than in the pre-test (Table 32). The significant difference, however, was not retained in the delayed posttest although the number of clauses per T-unit in the delayed posttest was higher than that in the pre-test (Pre-test Mean–Posttest2 Mean=-.072,  $p=.093$ ). Repeated measures ANOVAs (Table 33) run to detect the change of each group individually revealed that the main time effect for NP was significant ( $p=.016$ ) and its scores in both the posttest ( $p=.011, d=.66$ ) and the delayed posttest ( $p=.008, d=.56$ ) were significantly higher than that in the pre-test with medium effect sizes. The changes of PTP and OLP over time were not significant.

In terms of the mean length of T-units (MLT), the learners' performance in this measure did not vary significantly from pre-test to the delayed post-test. None of the three groups improved significantly over time.

In short, the NP group improved significantly in one of the measures for syntactic complexity (i.e. C/T) over time. The PTP and the OLP group did not improve

significantly in either of the two measures.

### 6.3.3 Changes in lexical complexity over time

Two measures were used to assess lexical complexity of the texts: lexical variety (D) and lexical sophistication (Lam). The table below reports the descriptive statistics of the two measures.

Scores for both measures were normally distributed and thus were submitted to parametric tests. Repeated measures ANOVA tests (Table 35) did not find any significant time effect for lexical variety or lexical sophistication for any group.

Table 34 Descriptive statistics for lexical complexity in the three tests

Measures	PTP (N=25)		OLP (N=25)		NP (N=25)		
	M	SD	M	SD	M	SD	
Lexis							
	D						
	Pre-test	75.44	16.12	63.7	15.83	66.56	16.64
	Posttest 1	69.00	17.4	64.16	16.38	66.85	13.08
	Posttest 2	70.11	14.62	64.1	15.38	62.14	11.26
	Lam						
	Pre-test	1.09	.33	1.03	.36	.97	.22
	Posttest 1	1.09	.26	.96	.30	.92	.23
	Posttest 2	1.06	.26	1.03	.35	.85	.22

Table 35 Results for lexical complexity across tests

Measures	Group	MS	F	Sig.	Effect sizes (f)
Lam	PTP	.011	.13	.878	.07
	OLP	.038	.493	.614	.14
	NP	.097	2.598	.085	.33
D	PTP	295.929	1.75	.185	.27
	OLP	1.522	.012	.988	.00
	NP	173.895	1.041	.361	.21

### 6.3.4 Changes in fluency over time

Fluency was measured by words per minute (WPM) and percentage of self-corrections (Dysfluency). The descriptive statistics of these measures are displayed in the table below.

Table 36 Descriptive statistics for fluency in the three tests

Measures	PTP (N=25)		OP (N=25)		NP (N=25)		
	M	SD	M	SD	M	SD	
Fluency							
	WMP						
	Pre-test	10.00	1.22	9.92	1.93	9.06	1.77
	Posttest 1	8.39	1.82	8.56	2.37	8.97	2.05
	Posttest 2	7.67	1.31	9.27	2.51	8.27	1.80
	Dysfluency						
	Pre-test	3.49	2.62	3.62	2.46	4.21	3.20
	Posttest 1	3.18	2.22	4.05	2.83	4.68	3.48
	Posttest 2	2.74	1.69	3.86	2.42	5.00	3.70

The assumption of normal distribution was held tenable for WPM, thus a repeated measures ANOVA was conducted to detect statistically significant differences in this measure. Data on dysfluency was not normally distributed so non-parametric tests were used.

Table 37 Results for fluency measures across tests

Measure	Group	F	Sig.	Location of significance:		
				<i>p</i> value (effect sizes)		
				Pre-Post1 (d)	Post1-Post2 (d)	Pre-Post2 (d)
WPM	PTP	27.705**	<.001	<.001 (1.04)	.051 (.45)	<.001 (1.84)
	OLP	4.135*	.018	.002 (.63)	.548 (.30)	.164 (.28)
	NP	2.192	.123	----- (.04)	----- (.39)	----- (.44)
Dysfluency		Chi-square	Sig.	Pre-Post1 (r)	Post1-Post2 (r)	Pre-Post2 (r)
	PTP	3.20	.852	----- (.15)	----- (.19)	----- (.21)
	OLP	.080	.961	----- (.05)	----- (.11)	----- (.06)
	NP	1.68	.432	----- (.15)	----- (.01)	----- (.12)

Note: Dashes mean that the calculation was not conducted.

Repeated measures ANOVA results showed that the main time effect on WPM was significant ( $F(2,144)=16.452, p<.001$ ) and the effect size was medium ( $f=.48$ ). Pairwise comparisons revealed that learners' performance in this measure was significantly better in the pre-test than in both the post-test ( $p<.001, d=.43$ ) and the delayed post-test ( $p<.001, d=.67$ ) with small and medium effect sizes. The time by group interaction in this measure was also significant ( $F=4.614, p=.002$ ), indicating that the changes from the pre-test to posttest2 were not identical in the three groups. To further investigate the differences, repeated measures ANOVAs were run for each group separately. Results (Table 37)

showed that NP did not change significantly from the pre-test to the delayed post-test ( $F=2.192$ ,  $p=.123$ ) but OLP ( $F=4.135$ ,  $p=.018$ ) and PTP ( $F=27.705$ ,  $p<.001$ ) produced significantly fewer words per minute over time. Pairwise comparisons showed that OLP's performance in the post-test significantly deteriorated ( $p=.002$ ,  $d=.63$ ) with a medium effect size but the difference between pre-test and the delayed post-test was not significant ( $p=.164$ ). PTP wrote significantly slower in the post-test ( $p<.001$ ,  $d=1.04$ ) and the delayed post-test ( $p<.001$ ,  $d=1.84$ ) than in the pre-test with large effect sizes.

For the other measure for fluency, the percentage of self-repairs (dysfluency), the within-group differences were not statistically significant.

To summarize, the PTP's words per minute decreased significantly from the pre-test to posttest1 ( $p<.001$ ) and posttest2 ( $p<.001$ ) and the OLP's words per minute decreased significantly from the pre-test to posttest1 ( $p=.002$ ), while NP's words per minute did not decrease significantly over time. In terms of the percentage of self-corrections, learners' changes over time were not significant.

### **6.3.5 Changes in organization scores over time**

The organization of learners' written texts was measured by subjective scoring for text coherence and cohesion (Organization rating) and computer program generated scores for local cohesion and coherence (Coh\_Local) and global cohesion and coherence (Coh\_Global). The descriptive statistics of these measures are displayed in the table below.

Table 38 Descriptive statistics for organization in the tests

Measures	PTP (N=25)		OLP (N=25)		NP (N=25)	
	M	SD	M	SD	M	SD
Organization rating						
Pre-test	5.76	.77	5.22	.73	5.16	.72
Posttest 1	5.62	.63	5.27	.64	5.20	.65
Posttest 2	5.42	.62	5.29	.67	5.01	.70
Coh_Local						
Pre-test	.06	.98	.07	1.09	-.12	.96
Posttest 1	-.05	1.19	.01	.91	-.03	.88
Posttest 2	-.17	1.09	-.02	1.01	.22	.91
Coh_Global						
Pre-test	.14	.93	-.28	.94	.15	1.1
Posttest 1	.14	1.16	-.16	.68	-.06	1.05
Posttest 2	-.04	1.20	-.20	.98	.24	.79

Friedman's test for the subjective rating scores showed that the within-group difference was not significant for any of the three groups.

The same patterns were observed for computer-generated scores for local coherence and global coherence. The within-group differences were not significant for any group.

Table 39 Results for organization measures across tests

Measure	Group	MS	F	Sig.	Effect sizes (f)
Coh_Local	PTP	.315	.400	.673	.13
	OLP	.047	.070	.932	.05
	NP	.770	.989	.379	.20
Coh_Global	PTP	.255	.275	.761	.11
	OLP	.101	.138	.872	.08
	NP	.580	.654	.525	.17
Organization rating		df	Chi-square	Sig.	r
	PTP	2	4.816	.090	.30
	OLP	2	1.351	.509	.09
	NP	2	1.195	.550	.21

Note: Dashes mean that the calculation was not conducted.

### 6.3.6 Summary

Planning showed some positive effects over time, though limited, for accuracy. The PTP gained significantly in overall accuracy as measured by percentage of error-free clauses; the OLP became more accurate in the use of S-V agreement; and the NP did not

change significantly in accuracy.

There was no significant effect for complexity. In terms of syntactic complexity, the two planning groups did not change significantly in this aspect over time while the NP's number of clauses per T-unit increased from the pre-test to the posttest and the delayed posttest. With respect to lexical complexity, no group changed significantly in the two measures over time.

Regarding temporal fluency, both planning groups wrote significantly slower in the posttest than in the pre-test. The results for PTP were retained in the delayed posttest but those for OLP were not. NP did not change significantly in this aspect over time. No group had significant change over time in dysfluency.

With respect to organization, there was no significant change in any group.

#### **6.4 Between-group differences**

To compare the three groups' performance in the tests, ANOVA tests were conducted when the scores satisfied the assumption of normality of distribution and Kruskal-Wallis tests were used for scores that were not normally distributed. For the two measures (i.e. lexical diversity and organization rating) that showed significant between-group differences in the pre-test, gain score analyses (Jamieson, 2004) were used.



### 6.4.1 Between-group differences in accuracy

Table 40 Between-group differences in accuracy in the posttests

Measures	Chi-square	Sig.	Pairwise comparisons: <i>p</i> value (effect sizes)		
			PTP-NP (d)	OLP-NP (d)	PTP-OLP (d)
Comparative					
Posttest1	2.182	.336	---- (.33)	---- (.07)	---- (.39)
Posttest2	.746	.688	---- (.00)	---- (.22)	---- (.25)
S-V					
Posttest1	1.687	.43	---- (.18)	---- (.33)	---- (.22)
Posttest2	7.850	<b>.02</b>	.604 (.13)	<b>.005 (.84)</b>	.038 (.74)
%error-free					
Posttest1	.209	.901	---- (.08)	---- (.00)	---- (.08)
Posttest2	1.601	.449	---- (.00)	---- (.20)	---- (.23)
	F	Sig.			
Error/100					
Posttest1	.157	.855	---- (.00)	---- (.00)	---- (.00)
Posttest2	.192	.825	---- (.00)	---- (.18)	---- (.18)

Note: Dashes mean that the tests were not run.

Results showed that in the correct use of comparative forms there were no significant between-group differences in the post-test ( $H(2)=2.182$ ,  $p=.336$ ) or in the delayed post-test ( $H(2)=.746$ ,  $p=.688$ ). As for the correct use of S-V agreement, there was no significant group difference in the post-test ( $H(2)=1.687$ ,  $p=.43$ ). However, there was a significant between-group difference in the delayed post-test ( $H(2)=7.85$ ,  $p=.02$ ). To further analyze the difference, three Mann-Whitney tests were run. A Bonferroni correction was applied and the effects were reported at a .0167 ( $0.5/3=.0167$ ) level of significance. In the delayed posttest, the NP scored significantly lower than the OLP in the correct use of S-V agreement ( $U=189$ ,  $p=.005<.0167$ ,  $d=.84$ ). The effect size was large. The difference between the NP and the PTP ( $U=288.5$ ,  $p=.604$ ,  $d=.13$ ) and that between the PTP and the OLP were not statistically significant ( $U=228.5$ ,  $p=.038$ ,  $d=.74$ ).

With regard to the two general measures, there was no significant between-group difference in them in either the post-test or the delayed post-test.

In summary, the only between-group difference was found in the correct use of S-V agreement in the delayed posttest, with the OLP outperforming the NP.

### 6.4.2 Between-group differences in syntactic complexity

In the two measures for syntactic complexity, there was no significant between-group difference in either the post-test (MLT:  $H(2)=.412$ ,  $p=.813$ ; C/T:  $F(2,72)=.447$ ,  $p=.642$ ) or the delayed post-test (MLT:  $H(2)=.302$ ,  $p=.86$ ; C/T:  $F(2,72)=.744$ ,  $p=.479$ ).

Table 41 Between-group differences in syntactic complexity in the posttests

Measures	Chi-square	Sig.	Pairwise comparisons: <i>p</i> value (effect sizes)		
			PTP-NP (d)	OLP-NP (d)	PTP-OLP (d)
MLT					
Posttest1	.412	.813	----- (.02)	----- (.13)	----- (.19)
Posttest2	.302	.860	----- (.03)	----- (.20)	----- (.15)
	F	Sig.			
C/T					
Posttest1	.447	.642	----- (.18)	----- (.25)	----- (.06)
Posttest2	.744	.479	----- (.34)	----- (.20)	----- (.16)

Note: Dashes mean that the tests were not run.

### 6.4.3 Between-group differences in lexical complexity

Data on both measures for lexical complexity was normally distributed and thus was submitted to parametric tests. Since there was a significant between-group difference in D in the pre-test, the gain scores of the three groups were compared.

Analysis of variance found that there was a significant difference between groups in lexical sophistication ( $F(2,72)=4.784$ ,  $p=.011$ ). Post hoc analysis found that PTP performed significantly better than NP in this measure ( $p=.008$ ) while the differences between OLP and NP or PTP and OLP were not significant. To further investigate the group difference, two one-way ANOVA tests (Table 42) were conducted and revealed that the groups did not perform significantly differently in the post-test ( $F(2,72)=2.755$ ,  $p=.07$ ) but the difference between groups was significant in the delayed post-test ( $F(2,72)=4.139$ ,  $p=.02$ ). Post hoc tests showed that PTP had a significantly higher score in Lam than NP in the delayed post-test ( $p=.03$ ) and the effect size was large ( $d=.87$ ).

Table 42 Between-group differences in lexical complexity in the posttests

	F	Sig	Pairwise comparisons: p value (effect sizes)		
			PTP-NP(d)	OP-NP(d)	PTP-OP(d)
Lam					
Posttest1	2.755	.07	---- (.69)	---- (.15)	---- (.46)
Posttest2	4.139*	<b>.02</b>	<b>.03 (.87)</b>	.073 (.62)	1.00 (.09)
D					
Posttest1	.989	.377	---- (.32)	---- (.009)	---- (.36)
Posttest2	.805	.451	---- (.05)	---- (.30)	---- (.32)

Note: Dashes mean that the calculation was not conducted.

For lexical diversity (D), no significant between-group effects were detected.

In sum, in measures for complexity, the only significant between-group difference was in lexical sophistication in the delayed posttest, with PTP outperforming NP.

#### 6.4.4 Between-group differences in fluency

Fluency was measured by words per minute (WPM) and percentage of self-corrections (Dysfluency). Results (Table 43) showed that there was no significant between-group difference in the post-test ( $F=.502$ ,  $p=.607$ ) but the between group difference was significant in the delayed post-test ( $F=4.46$ ,  $p=.015$ ). Post hoc analysis found that OLP produced significantly more words per minute than PTP in the delayed post-test ( $p=.013$ ) and the effect size was large ( $d=.80$ ).

For the other measure of fluency, the ratio of self-repairs, there was no significant between-group difference in the post-test ( $H(2)=2.03$ ,  $p=.362$ ) but the between group difference was significant in the delayed post-test ( $H(2)=6.609$ ,  $p=.037$ ). To analyze further the difference, a series of Mann-Whitney tests (Table 43) were run and the significance level was adjusted by dividing .05 by 3 ( $.05/3=.0167$ ). The test results showed that PTP made significantly fewer self-corrections than NP ( $U=-2.513$ ,  $p=.012<.0167$ ,  $d=.79$ ) with a medium effect size while the differences between OLP and NP ( $p=.295$ ) and that between PTP and OP ( $p=.118$ ) were not significant.

To summarize, significant between-group differences were found in both fluency measures in the delayed post-test. The OLP wrote significantly faster than the PTP group and the PTP group made significantly fewer self-corrections than the NP group.

Table 43 Between-group differences in fluency in the posttests

	F	Sig	Pairwise comparisons: p value (effect sizes)		
			PTP-NP (d)	OLP-NP (d)	PTP-OLP(d)
<b>WPM</b>					
Posttest1	.502	.607	----- (.30)	----- (.19)	----- (.08)
Posttest2	4.46*	<b>.015</b>	.824 (.38)	.203 (.47)	<b>.013 (.80)</b>
	Chi-square	Sig.			
<b>Dysfluency</b>					
Posttest1	2.03	.362	----- (.51)	----- (.20)	----- (.34)
Posttest2	6.609*	<b>.037</b>	<b>.012 (.79)</b>	.295 (.36)	.118 (.54)

Note: Dashes mean that the tests were not run.

#### 6.4.5 Between-group differences in organization

The organization of learners' written texts was measured by subjective rater scoring for text coherence and cohesion (Organization rating) and computer program generated scores for local cohesion and coherence (Coh\_Local) and global cohesion and coherence (Coh\_Global). Considering that there was a significant between-group difference in Organization rating in the pre-test, the three groups' gain scores in this measure were compared.

ANOVA results (Table 44) showed that there was no significant between-group difference in the subjective rating scores in the post-test ( $H(2)=5.617$ ,  $p=.06$ ) or in the delayed post-test ( $H(2)=3.948$ ,  $p=.14$ ).

The same patterns were observed for computer-generated scores for local coherence and global coherence. There were no significant between-group differences in local coherence ( $F(2,71)=.101$ ,  $p=.904$ ) or global coherence ( $F(2,71)= 1.87$ ,  $p=.162$ ).

Table 44 Between-group differences in organization in the posttests

	F	Sig.	Pairwise comparisons: p value (effect sizes)		
			PTP-NP(d)	OLP-NP(d)	PTP-OLP(d)
Coh_Local					
Posttest1	.016	.985	---- (.04)	----- (.05)	----- (.00)
Posttest2	1.103	.337	---- (.42)	----- (.24)	----- (.18)
Coh_Global					
Posttest1	.908	.408	---- (.24)	----- (.11)	----- (.38)
Posttest2	1.249	.293	---- (.28)	----- (.49)	----- (.14)
	Chi-square	Sig.			
Organization					
Rating					
Posttest1	5.617	.06	---- (.66)	----- (.11)	----- (.55)
Posttest2	3.948	.14	---- (.62)	----- (.41)	----- (.20)

Note: Dashes mean that the calculation was not conducted.

## 6.5 Summary of results

Table 45 displays a summary of significant between-group differences and within-group differences in the posttest and the delayed posttest.

Table 45 Summary of significant between-group and within group differences in the posttests

Measures	Tests	Time	Group
Accuracy			
Specific measures	Posttest1		
Comparative	Posttest2		
S-V	Posttest1 Posttest2	OLP: Pre<Post2	OLP>NP
General measures	Posttest1		
Error/100	Posttest2		
%error-free	Posttest1 Posttest2	PTP: Pre<Post2	
Complexity			
Syntactic	Posttest1	NP: Pre<Post1	
C/T	Posttest2	NP: Pre<Post2	
MLT	Posttest1 Posttest2		
Lexical			
D	Posttest1 Posttest2		
Lam	Posttest1 Posttest2		PTP>NP
Fluency			
WPM	Posttest1 Posttest2	PTP, OLP: Pre>Post1 PTP: Pre> Post2	OLP>PTP
Dysfleuency	Posttest1 Posttest2		PTP>NP
Organization			
Organization rating	Posttest1 Posttest2		
Coh_Local	Posttest1 Posttest2		
Coh_Global	Posttest1 Posttest2		

To summarize, for both the PTP group and the OLP group there were some significant gains in accuracy over a 14-week period. PTP's ratio of error-free clauses was significantly higher in the delayed posttest than the pre-test and OLP's use of S-V agreement was significantly better in the delayed posttest than in the pre-test. These results indicate that planning had some, but limited, positive effects on accuracy in the long run. With regard to complexity, neither the PTP nor the OLP had significant gains

over time. The NP group improved significantly from the pre-test to both posttests in one of the syntactic measures (i.e. C/T). In terms of fluency, planning had negative effects on temporal fluency for both planning groups. However, the effects were sustained at the delayed posttest as well as the posttest for the PTP only. None of the groups improved in organization over time.

Comparisons of performance of the three groups revealed that there were no significant between-group differences in all the measures in the post-test. However, there were significant between-group differences in the correct use of S-V agreement, lexical sophistication (Lam), ratio of corrections (Dysfluency), and number of words per minute (WPM) in the delayed posttest, indicating some longer-term effects of planning. The OLP outperformed the NP in the correct use of S-V agreement. The PTP produced significantly more lexically sophisticated words than the NP and it made significantly fewer self-corrections than the NP. The OLP produced significantly more words per minute than the PTP in the delayed post-test.

## **6.6 Discussion**

In this section, the effects of pre-task planning and online planning will be discussed separately. Since no previous studies have been found, to date, to investigate the longer term effects of these two types of planning on L2 writing, it is not possible to compare the results of the present study with other studies.

### **6.6.1 Longer-term effects of pre-task planning**

The within-group difference of PTP in the percentage of error-free clauses ( $p=.05$ ) was significant.

Seen from these results, it seems that repeated practice under the pre-task planning condition had some effects on the overall accuracy of written production. There are two possible reasons for the results: one has to do with the condition in which the pre-task planners had been practicing and the other has to do with the condition in which the pre-task planners took the posttests.

Regarding the first reason, pre-task planning seemed to have induced attention to accuracy. As will be seen in Chapter 7, fifty percent of the pre-task planners reported in the post-task interview having paid attention to accuracy while writing, though they did not receive specific instruction on how to plan the language. This percentage is higher than that of both the online planning group (23%) and the no planning group (24%). Therefore, it seems that pre-task planning offered more opportunities to attend to accuracy during production. Though this attention to accuracy did not lead to immediate significant effects on accuracy as has been reported in Chapter 5, it is possible that it helped learners gain more control of L2 in the long run. As Ortega (1999) explained, IL development proceeds via small changes and the benefits of attending to form might not have an immediate payback. However, as Ortega put it, “there is still possibility that in providing the space for sustained devotion of conscious attention to formal and systematic aspects of the language required by a task, planning yields benefits that are indirect and cumulative” (p.138).

The result can be explained by Skill Acquisition Theory and Skehan’s model of interlanguage development discussed in Chapter 2. According to DeKeyser’s account of skill acquisition (DeKeyser, 2007c), language skills development goes through three stages: declarative, procedural, and automatic. In the process of proceduralization, L2 knowledge is automatized, which could be then used more spontaneously. Automatization is a gradual process and progress in this process is characterized by reduction in reaction time (i.e. increased processing speed) and error rate. In the present study, it seems that pre-task planning provides a favorable condition for learners to advance on the way to automatization. It also corresponds to the second stage (gaining control of the L2 system) in Skehan’s model of IL development. The advancement in automatization or control of L2 was shown in reduction in error rate in the present study.

The second possible reason for the gains in overall accuracy of the pre-task planners might have to do with the condition in which the pre-task planners carried out the tasks in the testing sessions, which was different from the condition in which they performed the tasks in the treatment sessions. In the treatment sessions, the PTP had ten minutes for



planning and fifteen minutes for completing the tasks, whereas in the testing sessions they had 25 minutes to complete the tasks without pre-task planning time. Therefore, in the testing sessions, they had more opportunities to engage in online monitoring. As has been discussed earlier, pre-task planning had induced their attention to accuracy. Some pre-task planners may have established a production habit of attending to accuracy through repeated practice of writing under a pre-task planning condition. If this was true, then the habit of attending to accuracy together with more processing time may have caused the increased overall accuracy in the posttests. However, since how students went about doing the tasks in the testing sessions was not investigated, the above explanation is just a hypothesis, which needs future research to verify.

However, there was no significant gain in the specific measures for accuracy. As claimed in Skill Acquisition Theory, though practice can facilitate automatization, we cannot expect the same level of accuracy and fluency for different structures at any point of time (DeKeyser, 2007a). It is possible that the use of the two structures were not further developed as other structures, which were not specifically measured in the present study. This might be the case particularly when no special attention was given to the two structures. No instruction was given to pre-task planners to pay attention to the two structures and post-study questionnaire results revealed that nobody in PTP was aware that the two structures were targeted in the study. Moreover, in the treatment sessions the PTP completed the tasks under time pressure. Unlike the online planners, most pre-task planners did not have time to edit their essays (see Chapter 7 Section 7.4) and their awareness of the problematic areas, such as S-V agreement and comparative forms, might not be raised through practice. Therefore, even when having the opportunities to monitor their output in the testing sessions they might not attend to the target structures specifically. Moreover, the limitation with these specific measures discussed in Chapter 5 (i.e. they were already performing at a high accuracy level at the onset of the study) made it harder for this group to achieve significant gains. The results for the specific measures and general measures for accuracy suggest that attending to accuracy is facilitative to automatization but the same level of development in different structures should not be

expected.

With regard to *complexity*, test results show that PTP's within-group differences were not significant in either of the syntactic complexity measures while NP's within-group difference in number of clauses per T-unit was statistically significant ( $p=.023$ ). No significant between-group difference could be found in either the post-test or the delayed post-test. The result for PTP is not surprising since in the treatment sessions the pre-task planning condition did not show any advantage in this aspect.

It has been discussed in Chapter 5 that a feature of the tasks in the present study (i.e. the provision of ideas that participants could draw on when completing the tasks) reduced the necessity for participants to generate their own ideas, which in turn might have weakened any immediate advantage for syntactic complexity found in other research (e.g. R. Ellis & Yuan, 2004; Ojima, 2006) from the opportunities to plan the propositional content before the task. Thus, it is not surprising that repeated practice on writing tasks with such a feature, in spite of the opportunity to plan prior to task, did not lead to significant gains in syntactic complexity over time. This, to some extent, supports Robinson's (2003, 2005; Robinson & Gilabert, 2007) claim that reducing task complexity along the resource-dispersing dimension alone might not be conducive to L2 development. According to him, increasing resource-dispersing variables would facilitate access to the existing L2 knowledge, while increasing resource-directing variables would push learners to extend their L2 repertoire. Synergies of the two dimensions should be pursued in task design to help develop interlanguage because resource-dispersing variables (e.g. provision of planning time and background information) can free up some of learners' attentional and memory resources for allocation to the resource-directing dimensions of tasks. Tasks in the present study might have too many resource-dispersing variables (e.g. provision of planning time, background information, and task structure) and too few resource-directing variables (e.g. requirement for use of first-person, second-person, and third-person perspective). This might have affected L2 development in terms of complexity.

In addition, it has also been discussed in Chapter 5 that there are problems with the

measures for syntactic complexity used in the present study, which are popular in planning studies. Some studies (e.g. Knoch, 2007; Lu, 2011) have shown that number of clauses per T-unit is a poor index for syntactic complexity because it cannot differentiate essays produced by learners of different proficiency levels. Biber et al. (2011) argued that clausal subordination and T-unit analysis might be more appropriate for analyzing oral production than academic written production. It could be because of these problems that the results for syntactic complexity of the present study were perplexing (i.e. non-significant within-group difference for the pre-task planning group but a significant within-group difference for the no planning group).

The finding that NP significantly improved in C/T should be interpreted in relation to its decreasing scores of lexical sophistication (will be further discussed later in this section). As Skehan's study (Skehan, 2009a) indicated, there was a negative correlation between lexical sophistication and syntactic complexity. The growing reliance of NP on more common words in writing may have provided more space for NP to produce more structurally complex sentences. Since studies that used lexical sophistication to measure task performance were still scanty, Skehan's finding needs further support and the above hypothesis awaits to be confirmed by future studies.

Seen from the results for syntactic complexity, it seems that pre-task planning has no effect on L2 writing complexity. To push learners to use the upper limit of their interlanguage, tasks of greater conceptual complexity, which provide more space for the pre-task planners to take advantage of pre-task planning time, might have to be used. More appropriate measures for L2 writing complexity need to be developed.

Regarding lexical complexity, there were no within-group differences in either lexical variety (D) or lexical sophistication (Lam). However, there was a significant difference in Lam between PTP and NP in the delayed post-test ( $p=.003$ ). This result is also consistent with results from the treatment sessions, in which PTP significantly outperformed NP in this aspect. An examination of the mean scores (see Section 6.3.3 Table 34) found that the significant difference in the delayed post-test was caused by NP's decreased scores in this measure. It was possible that with minimum time to do

strategic planning and online planning NP had to use the words they were most familiar with, which mostly consist of high frequency words, to complete the tasks, leading to significantly lower scores for lexical sophistication than PTP in the treatment sessions. The repeated practice of completing the tasks under a no planning condition might have caused the decrease in their scores for lexical sophistication. While the no planning condition had negative effects on lexical complexity, the pre-task planning condition did not seem to bring an increase in this respect. Although pre-task planning enabled learners to use significantly more sophisticated words in the treatment sessions than the no planning group, it did not raise this group's scores in the posttests. Pre-task planning lent opportunities for students to mobilize more advanced vocabulary (see Chapter 7 Section 7.3) but it could not expand students' bank of vocabulary. Production practice cannot lead to extension of existing lexical knowledge. Learners need input to add new words to their existing vocabulary bank. Therefore, it is not hard to understand why the within-group difference of the PTP in lexical sophistication was not significant.

Turning from complexity to *fluency*, results of the tests show that PTP's temporal fluency as measured by words per minute (WPM) decreased significantly from pre-test to the posttest and the delayed posttest, indicating a negative effect of pre-task planning on temporal fluency over time. One possible explanation for the decreasing writing speed of the PTP is that this group may have formed a habit of rehearsing the language (i.e. practice the language for subsequent writing) before putting it down on paper through the treatment. In the post-task questionnaire and interview, rehearsing (100% in the questionnaire and 79% in the interview) was one of the most reported planning strategies (see Chapter 7 section 7.3.2 for more detailed results). It shows that most of the pre-task planners made use of the planning time to rehearse the language prior to task completion. Moreover, it was found through an analysis of their planning notes that they had developed strategies (e.g. writing more complete sentences on note sheets) for applying the planned language in the subsequent task (see Chapter 7 section 7.4). Through repeated writing practice under the pre-task planning condition, they might have already got into a habit of rehearsing the language before writing, which might be carried over to the testing

sessions. In the testing sessions, they did not have a fixed period provided for them to do pre-task planning as they did in the treatment sessions. However, they had more time (25 minutes) than in the treatment sessions (15 minutes) for completing the tasks, which allowed them to allocate some time to rehearse the language before writing. Although in the testing sessions, they did not have a planning sheet on which they could jot down the language they might use in later writing, they could engage in mental rehearsal. It is not difficult to understand that rehearsing might have slowed down the writing speed but at the same time it might also have decreased the number of self-corrections, because after rehearsal the students would become more confident with the language they used. This is manifested in the decreasing ratio of corrections of this group over time, which caused a significant between-group difference in the delayed post-test ( $p=.012$ ) with the PTP outperforming the NP. It seems that pre-task planning led to reduction in temporal fluency but improvement in repair fluency.

This finding does not seem to fit with the claim of Skill Acquisition Theory (i.e. automatization would lead to reduction in both reaction time, which is often measured as the time spent on giving an answer to a test item, and error rate) in that pre-task planning led to an increase in overall accuracy but not in temporal fluency. It needs to be noted, however, temporal fluency measured in the present study could not truly indicate reaction time. In writing, how fast a writer responds to the input materials may not be measured by the speed of writing. A writer with quick reaction may choose to write slowly for various reasons. For example, s/he may prioritize neatness of handwriting over speed of writing, which might be the case of the present study. Therefore, the result for temporal fluency is not really incompatible with the result for accuracy.

With regard to *organization* of the essays, there were neither significant between-group differences nor within-group differences in organization measures. As seen from the results for organization in the experimental sessions and the testing sessions, pre-task planning did not have any effect on text organization. Task features and students' limited discourse knowledge of English essays could be the two main reasons for this result. As has been discussed in Chapter 5, for each task used in the present study

some information on the objects under comparison was provided on the task sheets either in the form of tables or bulletin lists. This seemed to have provided a structure for the task for some participants, though this was not intended. They basically followed the order of the provided information in structuring the essays, therefore, producing texts with similar structures. Another consideration was that participants had limited discourse knowledge of English essays in general and knowledge of compare/contrast essays in particular. At the time data for the present study was collected, they had not been taught formally how to write this type of essay (see Chapter 5 Section 5.6.5.3). Therefore, their difficulty with constructing cohesive and coherent essays could not be overcome by planning opportunities.

To sum up, the opportunity offered by pre-task planning to attend to accuracy seems to have a cumulative effect on overall accuracy in a longer term, suggesting some advancement in automatization of L2 knowledge. Provision of possible information for inclusion in the essays on the task sheets and the problems with measurement might have had an influence on the results obtained for syntactic complexity. Though pre-task planning had immediate positive effects on lexical complexity, it did not help learners develop in this aspect over time because planning opportunities could not enlarge their vocabulary size. With regard to fluency, it is hypothesized that the pre-task planners may have formed a habit of rehearsal from the repeated practice of writing under the pre-task planning condition and this habit had a negative effect on the temporal fluency of writing but a positive effect on repair fluency over time. Regarding the effects on organization, the unintentional provision of structure for the essays and students' limited discourse knowledge are the possible reasons for the lack of effect.

### **6.6.2 Longer-term effects of online planning**

To reiterate the results of the tests for *accuracy*, OLP had a significant gain from pre-test to the delayed posttest ( $p=.016$ ,  $r=.46$ ) in the correct use of S-V agreement and the difference between OLP and NP in this measure was also significant in the delayed posttest ( $p=.005$ ,  $d=.84$ ) with a large effect size. This indicates some effects of the online

planning condition on accuracy. The OLP had more time to complete the tasks than the other two groups in the treatment sessions. Part of the extra time was devoted to editing after they had drafted the essays. Nine out of the thirteen (69%) students interviewed in this group reported having time to edit the essays before they handed them in. Some errors were corrected during editing and six out of nine (67%) interviewees who had time to edit reported having corrected errors in the use of verbs, such as tense and third person singular 's' (see Chapter 7 Section 7.4.3 for details). When asked what kind of errors they were able to correct, one of the interviewees stated "well, errors in tense or third person singular 's'. I knew I often make errors in these forms so if I had time I would check these things." It is possible that identification of and correction to errors in S-V agreement raised learners' awareness of this type of errors. This awareness helped them reduce errors in S-V agreement in later production tasks, further automatization of this form. It seems that the awareness of the problematic linguistic areas raised during online monitoring and editing could increase the accuracy of the forms. The fact that students did not mention comparative forms as a structure they paid attention to may explain the lack of gains for this target structure. The above finding corroborates an argument made in the Output Hypothesis that in response to a gap that they have noticed through producing language, learners may analyze their existing linguistic resources, which might result in consolidation of their existing L2 knowledge (Swain, 1985, 1995). Therefore, it seems important to provide opportunities for learners to pay special attention to forms they have difficulty with to further automatize the partially proceduralized knowledge.

In measures for *complexity* (including both syntactic complexity and lexical complexity), no significant differences in complexity, within-group or between-group, were found for the online planning group. It seems that online planning did not have any impact on syntactic complexity or lexical complexity over time. As has been discussed in Chapter 5, although 54% of the participants in the OLP group reported having made efforts to use complex language when writing in the post-task interview, the limitations of their L2 knowledge seemed to have placed a ceiling on the outcome of their efforts. Some of the participants in this group complained in the interview that the number of sentence

patterns and vocabulary they could use were so limited that despite their efforts to use complex language they could not achieve the level of complexity they wished. Having more time to produce language alone could not help close the gap that had been noticed with students' existing L2 knowledge. Without further L2 assistance, learners might not be able to produce more complex language over time. Apart from the limitation of L2 knowledge, reasons proposed for the lack of effect for pre-task planning (task complexity and the problems with measurement) on syntactic complexity might also account for the lack of longer-term effect on complexity for online planning.

Results for *fluency* indicate that online planning has a negative effect on temporal fluency of writing. This group produced significantly fewer words in the posttest than in the pre-test ( $p=.002$ ,  $d=.63$ ) with a medium effect size. This could be explained by a transferred effect from the experimental sessions. In the experimental sessions, the OLP group wrote significantly slower than the other two groups (see Chapter 5 Section 5.4.4 for details), which indicates that online planning would result in a lower speed in writing. It is possible that the online planners did engage in more online planning than the other two groups in the experimental sessions as evidenced in the data from the post-task interview (see Chapter 7 for a detailed report and Chapter 5 for a more detailed discussion) and this habit of planning online might be carried over to the posttest, causing a decrease in temporal fluency. However, this negative effect on OLP was not sustained as its temporal fluency increased from the posttest ( $M=8.56$ ) to the delayed posttest ( $M=9.27$ ). This shows that the negative effect of online planning that was carried over to the posttest was weakened after ten weeks, when the delayed posttest was administered. During the ten-week interval, the habit they had developed through writing practice gradually died.

Online planning did not affect *organization* of the essays since no significant differences in measures for organization were found. The possible reasons accounting for the lack of effect of pre-task planning on organization discussed earlier might also apply to online planning. That is, the unintentional provision of essay structure and students' limited discourse knowledge might have influenced the results for organization of the



OLP.

### 6.6.3 Summary

Practicing under the two planning conditions repeatedly seemed to have some effects on accuracy over time. Pre-task planning seemed to have offered more opportunities for the students to attend to accuracy during writing. This had a cumulative effect and led to the significant progress of PTP in overall accuracy over time. The online planning condition lent more opportunities to editing, during which process students' existing knowledge about verb use was consolidated, thus leading to the significant gain of OLP in the correct use of S-V agreement over time.

The two planning conditions did not have benefit for complexity over time. Task features (e.g. provision of information for possible inclusion in the essays), problems with measures, and the students' limited L2 knowledge were argued to be the possible reasons for the lack of effect of task planning on syntactic complexity. Although pre-task planning produced immediate effects on lexical complexity, the repeated practice under the planning condition could not extend learners' existing lexicon, thus could not bring about a significant gain over time.

The fact that NP significantly increased in one of the two syntactic complexity measures was attributed to its weakening performance in lexical sophistication, which has been found to negatively correlate with syntactic complexity (Skehan, 2009a).

The two planning conditions decreased the temporal fluency over time but the pre-task planning condition helped students in the PTP make significantly fewer self-corrections than those in the NP. It is hypothesized that the habits of the two planning groups developed through repeated practice in the treatment sessions (i.e. rehearsal for the pre-task planners and monitoring and editing for the online planners) were carried over to the posttests and this in turn influenced the results for fluency.

With respect to the effects on organization over time, neither planning conditions showed any effect. It has been argued that the unintentional provision of essay structure and the students' limited discourse knowledge were the possible reasons for the lack of

effect on organization.

Based on the results outlined above, it can be concluded that the longer-term impact of task planning on L2 writing was limited, though there were some encouraging signs, such as increased accuracy in some measures for accuracy.



## CHAPTER7 THE PROCESS OF PLANNING<sup>15</sup> AND WRITING

### 7.1 Overview

This chapter addresses the following research questions:

Research Question 3: What do learners do during planning and writing?

- a. What do they focus on during pre-task planning time?
- b. What strategies do they use to plan their writing?
- c. How much of their plan is applied in subsequent writing?
- d. What are their planning styles?
- e. What processes do they engage in while writing the essays?
- f. How do they perceive their task completion conditions?

Research questions 3a, 3b, 3c, and 3d apply to the pre-task planning group only whereas question 3e and 3f applies to all participants. The chapter begins with an overview of data analysis (section 7.2) and then presents results for the questions listed above. Results for RQ3a and RQ3b will be presented in section 7.3 titled ‘Process of planning’. Section 7.4 will report results for RQ3c and RQ3d. After that, what students focused on during writing and the strategies they employed (RQ3e) will be reported (section 7.5), following by participants’ perception of their respective task completion conditions (RQ3f) in section 7.6. Then, in the last section of the chapter (section 7.7) results for RQ3 will be discussed in relation to previous studies on the process of planning and to the effects of planning on L2 writing.

### 7.2 Overview of data analysis

To answer Research Question 3, a post-task questionnaire and a semi-structured interview were administered (see Appendix E & F). All participants filled out the questionnaire and half of them took part in the interview (see Chapter 4 Section 4.7). The table below displays a summary of data used to answer Research Question 3.

---

<sup>15</sup>In this chapter, the word ‘planning’ is used as shorthand for pre-task planning.

Table 46 Summary of data used to answer Research Question 3

Research question	Data
3a. Focus of planning	(for PTP) Questionnaire Q3; Interview Q4
3b. Planning strategies	(for PTP) Questionnaire Q1; Interview Q3
3c. Application of plan	(for PTP) Questionnaire Q6; Interview Q8; Planning notes
3d. Planning styles	(for PTP) Planning notes
3e. While-writing processes	Interview Q10-17 for PTP; Q1-10 for OLP and NP
3f. Learners' perception	Questionnaire Q12&13 for PTP; 10&11 for OLP and NP Interview Q18&20 for PTP; 12&13 for OLP and NP

For RQ3a, a conventional approach of content analysis (see Chapter 4 section 4.8.2) was conducted and coding categories were derived directly from data. For RQ3b and 3e, Ortega (2005) and Guar-Tavares's (2008) categories of planning strategies were used as reference. RQ3c was answered by participants' answers to questions in the questionnaire and the interview and verified by their planning notes. Their planning notes were also used to answer RQ3d. RQ3f includes answers from all participants about their perceptions of their respective task completion conditions. PTP's comments were coded with reference to Ortega's (2005) classification of reported benefits of planning with adaptations. OLP and NP's comments were categorized according to themes and patterns derived from data. A more detailed description of coding categories for each question will be presented in the following sections.

### 7.3 The process of pre-task planning

This section will cover the focus of planning and planning strategies. A general picture of the whole group will be presented by the results of the questionnaire analysis since all participants completed the questionnaire. The results of the interview data analysis will be reported as a complement or verification of questionnaire results.

#### 7.3.1 Focus of planning

In both the post-task questionnaire and the interview, the students were asked to answer what they focused on during planning with an open-ended question to answer RQ3a (Questionnaire Question 3&Interview Q4: What would you say your focus was

when you were preparing to write?). In the questionnaire, they provided brief answers to the question, such as “content of the essay”, “content, vocabulary”, and “structure of the essay, grammar”, whereas in the interview they provided their answers with a bit more details. For example, one student reported that “I focused on the *structure* of the essay because I think structure is quite important. Without a clear structure the essay will be messy.” According to their answers, the reported foci of planning were classified into three broad categories: *structure*, *content*, and *language*. These categories cover answers that contained explicit mention of the words (i.e. *structure*, *content*, and *language*) and those that clearly indicated that the student was focusing on one of these aspects. For example, a comment like “most of the time I was thinking about the points I should include in my essay” was categorized as *content* and “I was focusing on thinking about the words I could use” would be considered as planning on *language*. It is acknowledged that the three aspects are interwoven. Structure and language cannot stand alone without content. In this thesis, the three categories should be understood as the following: the *content* category refers to attention paid to idea generation and idea selection; *structure* refers to the particular attention paid to the organization of the content; and the *language* category emphasizes the linguistic encoding of ideas. It needs to be noted that the above definitions are given by the researcher. Whether these terms were understood in exactly the same way by the participants was not guaranteed.

The aspect that was mentioned the most in response to the question in the questionnaire was *content* (19 times), followed by *language* (14 times), and then *structure* (11 times). It should be noted that there was not only one focus for participants. Eight students (32%) reported both *content* and *language* being their focus of planning. Seven students (28%) indicated that both *structure* and *content* were their foci and five students (20%) focused on both *structure* and *language*. Table 47 shows the raw number of times each aspect was reported and the percentage of participants who reported each of the aspects. In this table *vocabulary*, *sentence patterns*, and *grammar* are listed as sub-categories of *language*. It seems that more students focused on planning the *content* (76%) and *language* (56%) than the *structure* (44%) of the essay. In terms of *language*,

more students reported planning *vocabulary* (40%) than those who reported planning *sentence patterns* (12%) and *grammar* (16%).

Table 47 Questionnaire results for focus of planning by pre-task planners

Aspects	Raw	Percentage (n=25)
Content	19	76%
Structure	11	44%
Language	14	56%
Vocabulary	10	40%
Sentence pattern	3	12%
Grammar	4	16%

Note: Raw=raw number of times the aspect was mentioned; Percentage=percentage of participants

The answers from the interview were consistent with the above results. Among the 14 students interviewed, eight (57%) reported that they focused on working on the *content* of the essays, eight (57%) reported having focused on *language*, and three (12%) mentioned that they paid most of their attention to *structuring* their essays. Of the eight students who reported having focused on *language*, two students (25%) reported having spent most of the time thinking of the sentences to use and six (75%) were searching for vocabulary most of the time. It should be highlighted that there was not only one focus for each student. Seven students (50%) reported having focused on both *content* and *language*. Two students (14%) focused on both *structure* and *content*. Moreover, a check on answers of those who completed both the questionnaire and the interview found that these students provided consistent information, verifying the consistency and reliability of these data.

From the answers given in the questionnaire and the interview, a pattern could be observed. That is, more students reported having focused on *content* and *language* than on *structure*. It should be noted that it is hard to disassociate content and language. Meaning and form are not dichotomous but interdependent. When preparing language necessary for meaning conveyance, the students were probably “simultaneously holding in long term memory the considerations regarding the message to be conveyed and the essential formal resources to convey it” (Ortega 2005, p.106). The inextricable relationship between meaning and form might be the reason why some students reported

having focused on both the propositional content and the language needed for meaning conveyance when planning. Another pattern that is observed is that when planning the language, more participants prepared vocabulary rather than sentence patterns or grammar for later use.

### **7.3.2 Planning strategies**

In the post-task questionnaire and the interview, the participants were asked to describe their planning process to answer RQ3b (Questionnaire Q1: Please describe in detail how you planned; Interview Q3: Could you please describe your planning procedure?). As has been described in Chapter 4, a content analysis was conducted on the answers to these questions to reveal their planning strategies. Ortega's (2005) and Guar-Tavares' (2008) categories of planning strategies were used to guide the coding process. However, the retrospective and concurrent protocols Ortega and Guar-Tavare analyzed were reflections of planning processes of oral production, thus the categories could not suit completely the need for the present study, which used written data. Therefore, categories that were particular to the speaking mode (e.g. selective listening, monitoring impact on listener) and categories that were not found in the data of the present study were excluded. In cases where a reported strategy did not find a counterpart in the priori categories (i.e. Ortega (2005) and Guar-Tavares (2008)), a new category was added.

Ortega's (2005) categories of planning strategies were adapted from O'Malley and Chamot (1990) and Oxford's (1990) taxonomy of learner strategies and Guar-Tavares' (2008) categories were drawn from only O'Malley and Chamot's framework (1990). In Oxford's and O'Malley and Chamot's works, metacognitive strategies are considered as higher order executive skills that entail planning (i.e. predicting the outcome), monitoring and evaluating the outcome, whereas cognitive strategies operate "directly on incoming information, manipulating it in ways that enhance learning" (O'Malley & Chamot, 1990, p.44). Following Ortega (2005) and Guar-Tavares (2008), the planning strategies were categorized into metacognitive strategies, which included organizational planning,



previewing, self-monitoring, and self-evaluation (O'Malley & Chamot, 1990); cognitive strategies, which included a wide range of strategies such as rehearsal, elaboration, and lexical compensation, etc., and social/affective strategies, which involved interaction with other people and control over affect (O'Malley & Chamot, 1990).

A summary of the planning strategies used by the pre-task planners in the present study with examples will be presented in Table 48. The underlined words indicate the part of the quote that determined coding for the particular strategy. The raw number of times each strategy was mentioned and the percentage of the participants reporting each strategy will be reported. In some cases, a strategy was reported by the same participant more than once in the interview. In these cases, the strategy was counted once for the sake of controlling for an over-representation of that strategy. In the table, the newly added categories will be marked with asterisks.

It is acknowledged that there were some overlap in the categories of planning strategies. *Writing as prompt/trial*, *rehearsing*, and *drafting* all involve writing words on the planning sheets. To distinguish these strategies, a decision was made to include under the category of *rehearsing* only reports that clearly indicated that the student was practicing for the subsequent task by practicing mentally and writing down words without writing a rough draft. *Writing as prompt/trial* was restricted to writing for the purpose of prompting, formulating thoughts or identifying problems. *Drafting* only refers to writing a rough draft.

The results of interview data will be presented together with the questionnaire results in Table 49.

Table 48 Summary of planning strategies reported in the present study

Strategies	Definition and Examples
<p>Metacognitive</p> <ol style="list-style-type: none"> <li>1. Previewing</li> <li>2. Organizational planning</li> <li>3. Monitoring</li> </ol>	<ol style="list-style-type: none"> <li>1. This refers to reading all the information on the task sheets to understand task requirements. E.g. S2: “<u>Every time I would first read the requirements of the task</u> and make a general plan.” (previewing followed by organizational planning)</li> <li>2. This involves making a general plan for the task, defining the main idea, and the structure of the essay. E.g. S4: “<u>I would form a mental outline, that is, what to write in the first paragraph, the second paragraph, and the third paragraph....</u>”</li> <li>3. This refers to monitoring of the written production. E.g. S8: “<u>I would write down some sentences and then found the places that needed revision.</u>”</li> </ol>
<p>Cognitive</p> <ol style="list-style-type: none"> <li>4. Writing as prompt/trial*</li> <li>5. Elaboration</li> <li>6. Rehearsing</li> <li>7. Avoidance</li> <li>8. Lexical compensation</li> <li>9. Translating</li> </ol>	<ol style="list-style-type: none"> <li>4. This refers to the strategy of writing down words or sentences selectively as prompts, trials, or an aid for thought formulation. E.g. S8: “<u>I would write down some simple sentences to remind me of what to write next.</u>” (prompt)</li> <li>5. This involves adding details to an idea or to the text. E.g. S11: “<u>After I translated my ideas, if I had time I would think of some more details to add in the essay.</u>”</li> <li>6. This refers to practicing for later writing. E.g. S6: “<u>I went through the points I was going to include in the essay.</u>”</li> <li>7. This refers to circumventing an idea because of a lack of vocabulary or grammar needed for expressing it. E.g. S13: “<u>I would first go through the whole thing on the task sheet (preview) and then select the points I am able to express. That is I would drop the aspects I felt difficult to write about.</u>” (avoidance)</li> <li>8. This refers to the strategy used to compensate for the failure of retrieving vocabulary or sentences they needed by, for example, substituting with other words or expressions. E.g. S3: “<u>Sometimes I really wanted to use a good word but I forgot, then I would use a very common word instead.</u>”</li> <li>9. This refers to translating from L1 to L2. E.g. S5: “<u>I used both Chinese and English when preparing. The influence of Chinese was unavoidable. Sometimes I thought in Chinese then translated the ideas into English.</u> But if the ideas can be expressed with words and expressions I am very familiar with I would use English directly.”</li> </ol>

<p>10. Lexical/syntactical search*<sup>16</sup></p> <p>11. Memorization*</p> <p>12. Grouping ideas*</p> <p>13. Drafting*</p>	<p>10. This refers to learners' search for vocabulary or sentences for expressing ideas. E.g. S14: "After I read the points on the task sheet (preview) <u>I thought of the sentence patterns and phrases or good words I may need to write about that.</u>"</p> <p>11. This refers to the effort of memorizing the points written down on the planning notes for later recall. E.g. S9: "<u>Before the planning note was taken away I would stop to carefully go through my notes and try to memorize as much as possible.</u>"</p> <p>12. This refers to learners' effort in trying to find relationships between points and putting related ideas together. E.g. S3: "<u>When I went through those points I found there were some relations between one point and another and I would naturally put them together.</u>"</p> <p>13. This refers to writing a draft of the essay using mostly complete sentences during planning time. E.g. S8: "I did not make an outline. <u>I started writing (the draft) right away.</u> In order to save time I wrote kind of illegibly and sometimes I used very simple expressions. <u>So I was able to write a draft.</u>"</p>
<p>Social/affective</p> <p>14. Seeking help*<sup>17</sup></p>	<p>14. This refers to learners' seeking help from external sources, such as dictionaries, classmates, or the researcher. E.g. S2: "<u>When I had difficulty coming up with a word I needed I occasionally asked my classmates for help.</u>"</p>

Note: 1. Categories are adapted from Ortega (2005) and Guará-Tavares (2008). An asterisk indicates a new category that emerged from the data. 2. The examples are taken from data of the present study. They were originally in Chinese and translated by the researcher.

Altogether nine strategies were reported in the questionnaire and fourteen strategies were reported in the interview. In the questionnaire, two metacognitive strategies (i.e. *previewing 48% and organizational planning 88%*) and one cognitive strategy (i.e. *rehearsing 100%*) were reported most frequently (with percentage close to or over 50%).

<sup>16</sup>In Guará-Tavares (2008) the category was termed 'lexical search'. In the present study the participants reported that they sometimes not only searched for lexical items but also sentence patterns, hence the term was changed to lexical/syntactical search.

<sup>17</sup>In Guará-Tavare (2008), the category was termed 'appeal for help', referring to learners' oral appeal for help. In the present study, the participants were completing writing tasks and sought help from either their classmates or dictionaries. Thus, 'seeking help' is thought to be a better term for the category.

Table 49 Planning strategies reported in the questionnaire and the interview

Strategies	<u>Questionnaire results</u>		<u>Interview results</u>	
	Raw	Percentage (n=25)	Raw	Percentage (n=14)
Metacognitive				
1. Previewing	12	48%	8	57%
2. Organizational planning	22	88%	13	92%
3. Monitoring	1	4%	1	7%
Cognitive				
4. Writing as prompt/trial*	3	12%	14	100%
5. Elaboration	5	20%	6	43%
6. Rehearsing	25	100%	11	79%
7. Avoidance	0	0%	2	14%
8. Lexical compensation	0	0%	4	29%
9. Translating	1	4%	7	50%
10. Lexical/syntactical search*	2	8%	9	64%
11. Memorization*	2	8%	4	29%
12. Grouping ideas*	0	0%	2	14%
13. Drafting*	0	0%	2	14%
Social/affective				
14. Seeking help	0	0%	1	7%

Note: Categories are adapted from Ortega (2005) and Guará-Tavares (2008). An asterisk indicates a new category that emerged from the data.

It should be noted that *rehearsing* covers both practice in content and in language and it includes both mental rehearsal and writing on planning sheets for the purpose of practicing. By rehearsing, a mental scaffold was built to help with application of the plan in the subsequent writing. The following excerpts exemplify the rehearsal operation the students engaged in.

S9 (Zheng): “I wrote down the words and phrases I wanted to use in the essays so I could remember them and apply them in my essays.” (rehearse language by writing on planning sheets)

S6 (Gao): “I first went through in my mind the points I was going to include in my essay and then thought about the words and sentences to use.” (rehearse content and language mentally)

In the interview, the three strategies (i.e. *previewing* 57%, *organizational planning* 92%, and *rehearsing* 79%) also enjoyed a high percentage of report. Besides, five more strategies were reported in the interview: avoidance, lexical compensation, grouping

ideas, drafting, and seeking help.

Three more strategies were also reported substantially more in the interview than in the questionnaire: *writing as prompt/trial* (100%), *translating* (50%), and *lexical and syntactic search* (64%). This could perhaps be attributed to the data elicitation methods. When doing the pen and paper questionnaire, the participants might have been reluctant to give more than simple and straightforward answers and there was no chance for the researcher to ask follow-up questions or questions for clarification. In the face-to-face interview, however, more data could be elicited because the researcher could ask follow-up questions, through which more information might be elicited. For example, in the case of *writing as prompt/trial*, when filling out the questionnaire, some students might have thought that they were required to describe their mental operations during planning and that the physical action of writing and the purpose of writing words on the task sheets were irrelevant or too obvious to mention. To be economical with words, they might have chosen not to mention it in their answers to the question. In contrast, in the face-to-face interview, they might not have felt a need for being economical with words and thus this strategy was reported more in the interview than in the questionnaire.

Among the three strategies that were mentioned more in the interview, *writing as prompt/trial* needs further explanation because the purposes of writing notes differed among the learners. The purposes were summarized into three categories: writing for prompting, writing for identifying problems (trial), and writing for thought formulation.

Six students (43%) reported writing for prompting. S14 (Liu) said, “I would put down some sentence structures and words to remind me of what to write next.” S8 (Zhou) expressed a similar idea: “The simple sentences would remind me of what to write next and this can help with the logic and coherence of the essay.”

Writing on the task sheets was also reported to be helpful for identifying problems both in language and in content. S6 (Gao) expressed in the interview, “if you just think about it in your mind you won’t find the problems. But if you write it down during preparation you may find that some of your points are problematic then you can amend it when you write the essay.” S8’s words show how writing during planning could help

identify language problems. She said “I also tried to write some complex sentences and if I found I could not write it correctly I would not use it in the essay.” The problem identification function of selective writing was mentioned by four of the students (29%).

The last purpose of writing during planning was to help learners formulate thoughts. S4 (Xue)’s words can express this well: “You will think more clearly if you write it down on paper. If you just keep it in your mind it will be a mess. If I put it down I will become clear about what to write and how to write.” Three interviewees (21%) reported writing for this purpose.

In summary, two most frequently used metacognitive strategies reported by the pre-task planners were *previewing* and *organizational planning*. Among the cognitive strategies, *rehearsing*, *writing as prompt/trial*, *translating*, and *lexical/syntactic search* were the most frequently reported.

#### **7.4 Plan application and planning styles**

To investigate how much of their plan was applied in the essays (RQ3c), the participants in the PTP group were asked to estimate the amount of plan that they managed to use in their subsequent writing in both the questionnaire and the interview (Questionnaire Q6; Interview Q8). Learners’ planning notes were analyzed to verify their responses (see below for information on how they were analyzed). It is acknowledged that what was written on the planning sheets could reflect only part of the students’ mental operations. Nevertheless, it provided objective data for analyzing what the students were planning, how much of their plan was used in subsequent writing, and the style of their planning. In the following parts, learners’ plan application will be reported first, followed by the planning styles (RQ3d).

##### **7.4.1 Plan application**

As has been described earlier, the question on how much of the pre-task plan was applied in the essays was answered by: 1) learners’ self-reports in the questionnaire and the interview, and 2) examination of learners’ planning notes. Q6 in the post-task

questionnaire asked “Did you use much of the content you had planned in your writing? Did you use much of the language you had planned in your writing?” Participants were asked to indicate their application of plan on a scale of 1--5, with 1 meaning ‘none’ and 5 meaning ‘all’. Table 50 shows the participants’ choices.

Table 50 Application of plans reported in the questionnaire

Aspect	Scale					Mean
	1 (none)	2	3	4	5 (all)	
Content	0	1	6	17	1	3.68
Language	0	14	11	0	0	2.44

Note: The total number of participants was 25.

The mean score for the estimated application of planned content was 3.68, which is higher than the score for application of language (M=2.44). This indicates that they thought they were able to use more planned content than language.

In the interview the participants were also asked to estimate the amount of plan they were able to use in the essay (Interview Q8). The estimated amount of planned content and language applied was reported to be around 80% ( $3.68/5=.75$ ) and 50% ( $2.44/5=.49$ ) respectively, which were consistent with their answers in the questionnaire.

To verify if the reported amount of plan applied in the essays was accurate, learners’ planning notes were compared with their essays by using Text\_Lex Compare, an online program from [www.lex tutor.ca](http://www.lex tutor.ca). The total number of words on the planning sheets and the percentage of the planned words that were used in the essays are presented in the table below.

Descriptive statistics of percentage of planned words applied in the essays confirms strategic planners’ verbal report on the amount of plans used in their subsequent writing. In Time 1 the percentage was 58% and it increased to 67% in Time 2, 71% in Time 3, and 70% in Time 4.

Table 51 Descriptive statistics of the total number of words on the planning sheets and the percentage of the planned words applied in the essays

Measures	T1		T2		T3		T4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total words	71.28	28.9	64.68	24.8	72.00	23.61	83.44	28.12
% planned words applied in the essays	.58	.12	.67	.11	.71	.09	.70	.09

Note: T1=Time 1 writing; T2=Time 2 writing; T3=Time 3 writing; T4=Time 4 writing.

To examine if there was any significant change in application of plans over time repeated measures ANOVA were run and the results are shown in Table 52.

Table 52 Repeated measures ANOVA results for total number of words on planning notes and percentage of planned words applied in the essays

Measure	F	Sig.	Cohen's f	Pairwise comparisons (p value)		
				T2 vs. T1	T3 vs. T1	T4 vs. T1
Total words	4.594*	<b>.005</b>	.44	.204	.942	.007
%planned words applied in the essays	10.368*	<b>&lt;.001</b>	.64	.095	<b>&lt;.001</b>	<b>&lt;.001</b>

As can be seen from the table, there were significant changes in both measures over time. The total number of words the students wrote down on their planning sheets increased significantly over time ( $F(3,72)=4.594$ ,  $p=.005$ ) and the effect size was large ( $f=.44$ ). It indicates that the learners were able to write more during planning time through practice.

As can be expected, the main time effect on the percentage of planned words applied in the essays was significant ( $F(3,75)=10.368$ ,  $p<.001$ ) with a large effect size ( $f=.64$ ). Pairwise comparisons revealed that significant differences started to show from Time 3. Time 2 was not significantly different from Time 1 ( $p=.095$ ) whereas both Time 3 ( $p<.001$ ) and Time 4 ( $p<.001$ ) were significantly different from Time 1. By Time 3 the percentage of planned words that were used in the essays was significantly higher than that of Time 1. As the students practiced more, they were able to use more of their planned words in their essays.

This indicates that, over time, they drew more on their memories in using the planned words to perform the tasks. They used pre-task planning as an opportunity to rehearse the language needed for task completion, which could explain why *rehearsal*



(79%) and *writing as prompt/trial* (100%) enjoyed a high percentage of report among planning strategies.

#### 7.4.2 Planning styles

Following Risemberg’s (1996) method of analyzing pre-writing plans, the planning styles were categorized into four types: simple listing of words and phrases (Type A), drafting, which means that the students were writing a rough draft of the essay (Type B), making connections between ideas by drawing lines between phrases or sentences or grouping ideas (Type C), and a mixture of Type A and B (Type A/B) (see Chapter 4Section 4.7.3 for details of the method of analysis).

The use of four types of planning styles is displayed in Table 53. No statistical test was conducted on this analysis because the sample size was too small for any appropriate test to produce meaningful results. Therefore, raw numbers were used in reporting the results. Type A (simple listing of words and phrases) generally decreased over time; Type A/B increased from Time 1 to Time 2 and 3 but decreased in Time 4; Type B (drafting) increased from Time 2 to Time 4 and Type C (making connections between ideas) remained more or less unchanged.

Table 53 Raw number of planning styles used in the treatment sessions

Time	Type A	Type A/B	Type B	Type C
1	8	6	4	5
2	4	13	4	4
3	1	10	10	4
4	2	5	13	5

The numbers show a general trend of moving from Type A planning to Type B. This indicates that as students practiced more, they wrote more complete sentences on their planning notes. Type C planning was the individual preference of certain students (in total only five students ever used this style of planning and four of them used this style all the time).

Type C reflects the strategy of *grouping ideas* and Type B corresponds to *drafting*. The result for Type C is consistent with students’ report of the strategy (*grouping ideas*: 14%) while the result for Type B is not (*drafting*: 14%). The discrepancy between the

result for Type B and the report on the *drafting* strategy shows the limitation of retrospective self-report. Sometimes there could be a gap between what the learners thought they were doing and what they actually did.

## **7.5 The process of writing for the tasks**

The writing process will be described in terms of allocation of writing time, while-writing focus and strategies employed by the learners during writing to answer RQ3e.

### **7.5.1 Allocation of writing time**

OLP and NP groups were not given any time to plan before they performed the task so it was necessary to find out how they allocated their writing time to understand their writing processes and to establish if the operationalization of the treatment conditions was successful, that is, to see if there were indeed differences as intended between the conditions in which the three groups carried out the tasks. Therefore, the participants in OLP and NP were asked to describe how they allocated their writing time in the interview (Appendix F, Q1: Please tell me how you made use of your writing time). In the following paragraphs the reports from OLP and NP will be presented separately.

A total number of 13 online planners (OLP) were interviewed. Eight out of 13 of them (62%) reported that they first spent a little time on reading the instructions, understanding the task requirements, making a general plan, three of them (23%) planned the structure of the essay, and one of them (8%) prepared some words and phrases before starting to write. The time 46% of them reported having spent on reading the instructions and making a general plan was about 2 to 3 minutes, though two of them estimated the time to be 5 minutes. Five of the students (38%) reported spending less than one minute to read the instructions and make a general plan before they started to write. This indicates that participants in this group spent some time engaging in pre-task planning to a varying degree. After reading the instructions and making a general plan, they started to write. Their while-writing strategies will be reported in the following sections.

In the 25 minutes allowed for completing the tasks, some students were able to edit their essays after writing. Nine of them (69%) reported having time to edit and the length of editing time ranged from 1 minute to 5 minutes. On average the reported editing time was 4 minutes. Five interviewees (38%) reported having no time to edit. They wrote till the time their essays were collected. The interviewees were also asked to estimate the time they spent on actual writing. Two participants (15%) said the time was about 15 minutes; two (15%) spent 17-18 minutes; three (24%) spent 20 minutes; four (31%) spent 21-23 minutes; and two (15%) used up the 25 minutes. The average estimated time was 21.3 minutes, which matches the time recorded by the researcher.

With regard to the no planning group (NP), a total number of 17 students from this group were interviewed and 16 interviewees (94%) reported that they spent one or two minutes reading task instructions, understanding task requirements, and making a general plan for the subsequent writing. Hence, the actual writing time was around 13 minutes. During these one or two minutes, nine of them (53%) made a rough plan for what they were going to write in the essays, two of them (12%) thought of the key words and sentences they would use in their essay, and one of them (6%) also thought of the register of the essay. Since they only had 15 minutes to write, most of them (94%) did not have any time to edit the essays. There was one student (6%), however, who reported having three or four minutes to edit after finishing the essays.

To recap, most of the participants in both OLP and NP groups spent a little time to read task instructions and make a general plan before they started to write. This indicated that the pre-task planning they engaged in was limited and the reported actual writing time for OLP (21.3 min on average) was longer than that for NP (about 13 min for most participants). Many participants in the OLP group had a little time to edit their essays before handing them in and some of them (38%) were able to correct some grammatical errors.

### **7.5.2 While-writing focus**

Information on what learners focused on while writing was gathered through the

post-task interview (Q9 for PTP, Q2 for OLP and NP: What do you think you were focusing on when you wrote?). As with the description on focus of planning, the reported focus of writing was coded into three broad aspects: content, structure, and language. Four sub-categories under language emerged from the verbal report: accuracy, complexity, vocabulary, and grammar. Table 54 below shows the while-writing focus of the three groups. Since participants could mention more than one focus, the percentages of the reported aspects do not add up to 100%.

Table 54 While-writing focus reported by participants

Focus	PTP (n=14)		OLP (n=13)		NP (n=17)	
	Raw	Percentage	Raw	Percentage	Raw	Percentage
Content	5	36%	7	54%	17	100%
Structure	0	0%	2	15%	0	0%
Language*	13	93%	13	100%	17	100%
Accuracy	7	50%	3	23%	4	24%
Complexity	3	21%	7	54%	5	29%
Vocabulary	3	21%	0	0%	0	0%
Grammar	0	0%	1	8%	1	6%

Note: \*Some students reported having focused on language in general without specifying which aspect of language they focused on. That is why the numbers of students reporting having focused on specific aspects of language do not add up to the number of students reporting a language focus.

As with the categories used in reporting the focus of planning, the category of *content* covers answers in which the word ‘content’ was explicitly mentioned (e.g. “I focused more on content”) and those that indicated a focus on selection of points to be included in the essays (e.g. “I was thinking which points I could put in the essay and what to say under each point.”). As for *structure*, answers in this category include direct mention of essay structure (e.g. “I paid a lot of attention to the structure of the essays”) and answers that indicate a focus on structure (e.g. “I thought about the ordering of points, like which points could be put together and which point should follow which so the essay looked organized”). In reporting a *language* focus, some students commented that they focused on linguistic encoding of the of ideas (e.g. “Most of the time I was thinking about how to express my ideas in English”) whereas some other students specified the aspect(s) of language they focused on (e.g. “I spent most of the time thinking about the vocabulary I needed”; “I made great effort to write complex sentences”; and “I paid a lot of attention

to accuracy”). Therefore, four sub-categories were created under the language category: accuracy, complexity, vocabulary, and grammar. Comments like “I paid a lot of attention to accuracy” belonged to the category of accuracy. Statements that indicated that the student paid particular attention to complexity of language were put in the category of complexity. For example, one student stated that “I spent a lot of time thinking how to make my sentences and words varied so my essay would be impressive.” Comments that indicated explicitly that the student spent most the time on lexical search would be in the *vocabulary* category (e.g. “I spent most of the time thinking of the words I can use in the essays.”) and those that showed that the student paid most of the attention to grammar would be in the *grammar* category (e.g. “I paid a lot of attention to grammar.”).

As is shown in the table above, all participants in NP reported having focused on both content (100%) and language (100%) during writing with some attention to accuracy (24%) and complexity (29%). As was reported by participants in this group, their priority was to complete the tasks in the time allowed. For example, one student commented that “After I went through the instructions and the information provided in the table, I would start to write. Because time was tight and there was a requirement on the length of the essay, I would think of what to write next and how to express the ideas as I was writing. I had to write fast so I could not pay particular attention to things like accuracy.” Another student said, “sometimes I could not find the word to express that meaning, so I just used one word with similar meaning. If it was not correct, it was just one error and it was okay.”

For OLP, all participants (100%) reported having focused on language and a little over half of them (54%) also focused on content. Two students reported having also focused on the structure of the essays. Moreover, more than half of the online-planners (54%) specified that they had paid attention to complexity of language. For example, one student reported that “I tried to write long sentences and use some complex sentences like attributive clauses. Long sentences could impress people.”

Participants in PTP, in comparison, mainly (93%) focused on language and fewer of them (36%) focused on content. Half of the pre-task planners (50%) reported that they

paid more attention to accuracy of language. As one student commented, “I paid more attention to accuracy because that was how we were trained in high school. My teacher in high school told me that a composition with accurate language was more likely to get a high mark in the college-entrance exam. So we have been trained to pay attention to accuracy.” The number of students in PTP (50%) reporting having paid attention to accuracy was larger than that of students in OLP (23%) and NP (24%). This is consistent with the answers they gave to one question in the questionnaire asking about if they attended to accuracy while writing (Q10 for PTP; Q7 for OLP and NP). For this question, the participants were asked to indicate on a five-point scale whether they attended to accuracy, with 1 meaning ‘no’ and 5 meaning ‘yes’. The average scores of the three groups were 3.06 for NP, 3.07 for OLP, and 3.59 for PTP.

The picture of PTP will be clearer when their reported while-writing focus is seen in relation to planning focus. To reiterate, the reported planning focus in the questionnaire was content (76%), language (56%), and structure (44%). During writing more attention was paid to language (93%) than content (36%) and structure (0%). It seems that much of the content planning, part of language planning, and part of structure planning were done during pre-task planning time. Thus, they were able to allocate less attention to planning the content and structure, leaving more space to meet the ‘moment-to-moment’ language requirements of communication. It is also noticeable that for while-writing focus 50% pre-task planners reported having focused on accuracy, 21% focused on complexity, and 21% on vocabulary, while for planning focus there was no mention of accuracy or complexity. This indicates that during planning students were focusing on retrieving and rehearsing the language, leaving accuracy and complexity to be dealt with till writing for the task.

In conclusion, almost all participants who were interviewed reported having focused on translating ideas into language during writing (i.e. coded as *language*, 93% for PTP, 100% for OLP and NP) and fewer pre-task planners focused on idea generation and idea selection (i.e. coded as *content*, 36% for PTP, 54% for OLP; 100% for NP) than the other two groups. During translation of ideas into words, 50% of the pre-task planners paid

attention to accuracy. 54% of the online planners made effort to use complex language. For the non-planners, the primary concern was completing the tasks. Fewer participants in this group had time to particularize either accuracy or complexity of language.

### **7.5.3 Strategies employed while writing**

Information on strategies used during writing was also mainly gathered through the interview (Q1-10 for the OLP and NP and Q10-17 for the PTP). One question in the questionnaire also elicited information on while-writing strategies (Q4 for PTP, Q2 for OLP and NP: Did you encounter any difficulties? What were they? How did you overcome them?). This question had the aim of eliciting whether planning influenced the strategies students use in dealing with writing difficulties.

An analysis of the above mentioned data found that the reported while-writing activities included many planning strategies, which supports the argument that planning, accompanied by partial translation of ideas, could occur at various stages of writing (Hayes & Nash, 1996; Kellogg, 1996). The results that have been reported in section 7.5.1 (students in OLP and NP spent a little time to make a pre-writing plan) also verified the above finding. Therefore, students' comments in the interview and the questionnaire were analyzed using the coding scheme for planning strategies (see Section 7.3) with some adaptations. That is, when a reported strategy did not find a counterpart in the categories for planning strategies, a new category was added. The reported strategies used during writing are summarized and displayed in the table below and the added categories are marked with asterisks.

Table 55 Reported strategies used during writing

Strategies	Results from the interview						Results from the questionnaire					
	PTP (n=14)		OLP (n=13)		NP (n=17)		PTP (n=25)		OLP (n=25)		NP (n=25)	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Organizational planning	0	0%	10	77%	11	65%	-----					
Previewing	0	0%	7	54%	16	94%	-----					
Monitoring	1	7%	1	8%	0	0%	-----					
Generating ideas*	0	0%	1	8%	3	18%	-----					
Improvising*	5	36%	11	85%	11	65%	-----					
Elaboration	5	36%	2	15%	4	24%	-----					
Avoidance	0	0%	2	15%	1	6%	1	4%	2	8%	2	8%
Lexi/syntactic compensation	2	14%	2	15%	3	18%	19	76%	8	32%	15	60%
Seeking help	1	7%	1	8%	0	0%	15	60%	15	60%	15	60%
Look up dictionary							10	40%	10	40%	8	32%
Ask classmates for help							5	20%	5	20%	7	28%
Lexi/syntactic search	7	50%	5	38%	2	12%	-----					
Grouping ideas	2	14%	1	8%	3	18%	-----					
Rereading*	0	0%	7	54%	0	0%	-----					
Pausing*	0	0%	9	69%	0	0%	-----					
Editing*	2	14%	9	69%	1	6%	-----					

Note: 1. Rereading=learners reread what they had just written; Pausing=learners paused before they finished writing; Editing=learners edited what they had written; Improvising=using whatever linguistic forms available to fulfill the moment-by-moment language requirements of communication.

2. The shaded part indicates strategies used to cope with writing difficulties.

3. Dashes mean data was not elicited.

Table 55 shows that PTP did not engage in metacognitive strategies during writing, such as *organizational planning* and *previewing*, as these had been dealt with during pre-task planning time. They did not have to generate ideas either after they had started writing the essays. However, more than half of the participants in OLP (77%) and NP (65%) had to do *organizational planning* before writing as they were not given any time to prepare before task. Almost all participants in NP (94%) reported having done *previewing* while 54% online planners reported having done this.

During writing, a strategy that most of the participants employed was *improvising*. That is, they used whatever linguistic forms available at the moment to translate their



ideas into words. It is not surprising that fewer students in PTP (36%) reported using this strategy than those in OLP (85%) and NP (65%), because part of the content and language needed for expressing the ideas had been planned before they started writing the essays. Because they had spent time planning content and language during planning time, more students in PTP were able to spend writing time elaborating their ideas (36%) than those in OLP (15%) and NP (24%). Another strategy that was reported more by PTP (50%) than OLP (38%) and NP (12%) was *lexical/syntactic search*. This indicates that more participants in PTP made efforts in searching for language that could better express their ideas.

Three strategies that were used more by OLP than the other two groups were *rereading, pausing, and editing*. Nine students (69%) in OLP stated that they paused either in the middle or at the end of a sentence whereas none of the participants in PTP and NP reported doing so. Most of the learners (eight out of nine) paused to search for words or sentence patterns they needed, one paused to think about grammar, and one paused to think of what to write next. Seven of the interviewees (54%) in OLP reported sometimes looking back at what they had written in the middle of writing to reread the points they had covered or to check grammar and again none of the participants in PTP and NP reported doing that. As online planning has been defined to include formulation processes (in terms of both content and language) and online monitoring (see Chapter 3 section 3.3), those pauses and rereading during writing could be regarded as indicators of online planning taking place in the course of writing. With regard to editing, nine (69%) interviewees in OLP reported having edited their essays before handing them in while much fewer students in PTP (two out of 14 or 14%) and in NP (1 out of 17 or 6%) were able to edit.

Since more than half of the interviewees had time to edit their essays, it was of interest to know what kinds of errors they corrected. Among the nine students who had time to edit, four indicated (31%) that they could not correct any error in their essays even though they were aware that there might be errors, because it was hard for them to identify errors in their own writing. The majority of the errors corrected by the other five

interviewees (38%) concerned verbs. Two people mentioned tense, two people mentioned singular/plural forms, one person mentioned verb forms, three people mentioned 3<sup>rd</sup> person 's', and one mentioned spelling. The students were aware that the use of verbs was a problematic area for them and thus they paid some attention to this when writing.

With respect to the strategies learners employed to cope with difficulties they encountered during writing, differences showed in their responses in the interview and those in the questionnaire. Higher percentages of participants reported using lexical compensation and seeking help strategies in the questionnaire than in the interview (see Table 55). This could be attributed to the fact that the participants were asked directly to answer how they coped with difficulties they had during writing in the questionnaire while in the interview the question was not asked but a follow-up question was asked when participants mentioned (a) difficult(ies) voluntarily. Below is an example of how a question on how they coped with difficulties was asked in the interview. The relevant question and answer are italicized.

Interviewer: So during writing most of the time your focus was on...

S5: thinking of the language to use to express my ideas.

Interviewer: Which aspect of language did you pay more attention to?

S5: I paid more attention to sentence patterns. Sometimes I would use some sentences for emphasis, such as "why not...". As for advanced words, sometimes I was afraid I could not use them correctly.

Interviewer: *In those cases, what did you do?*

S5: *It depends. Sometimes I would use another word. Sometimes I used another sentence.*

(lexical/syntactical compensation)

In general, to deal with language difficulties participants reported using *lexical/syntactical compensation* and *seeking help* more than they used *avoidance* strategy. More participants in PTP (76%) and NP (60%) reported using lexical compensation than those in OLP (32%). This could be due to the fact there was more time pressure for PTP and NP, as they had less time to complete the tasks than OLP. Participants in PTP and NP might have had to rely more on lexical compensation strategy

than OLP, who could afford more time to come up with the words they planned to use.

## **7.6 Learners' comments on task completion conditions**

Learners' comments on their respective task completion conditions were gathered through the post-task questionnaire (Q13&14 for PTP; Q10&11 for OLP and NP) and the interview (Q16&18 for PTP; Q10&11 for OLP and NP). Participants' answers to the questions will be reported in the following sections. One other question on the questionnaire was also relevant to task completion conditions. The question asked if they felt pressed for time (Q8 for PTP; Q5 for OLP and NP). The participants were asked to indicate on a five-point scale whether they felt pressed for time when writing with 1 meaning 'no' and 5 meaning 'yes'. The scores they gave were respectively PTP: 3.03, OLP: 2.7, NP: 3.47. Given the fact that OLP had 25 minutes and PTP and NP had 15 minutes to complete the tasks, it is not surprising that NP felt pressed for time the most and OLP felt the least pressed.

### **7.6.1 Learners' perceptions of pre-task planning**

In the questionnaire, the pre-task planners were asked to answer whether they thought having time to plan before writing was helpful by indicating on a five-point scale, with 1 meaning 'not helpful' and 5 meaning 'very helpful' (Appendix E Q13). The average score for this question was 3.5. They were also asked to write in what ways planning was helpful. In the interview, learners were asked to comment on the fact that they had 10 minutes to prepare before writing (Appendix F Q16&18). When coding the reported benefits of planning, categories summarized by Ortega (2005) were used as guidance. Adaptations were made to Ortega's classification. New categories, which are marked with asterisks in the table below, were added when reported benefits could not find their counterparts in Ortega's taxonomy.

Table 56 Summary of benefits of pre-task planning identified by learners

Benefits of pre-task planning	Examples
1. Help organize thoughts	1) My ideas can be better organized; 2) I could have time to make an outline.
2. Help formulate thoughts	It (planning) helped with coming up with the main idea and how to make arguments.
3. Help with structure*	It made the structure of the essays clear.
4. Improve overall content	I could think about the content more clearly.
5. Practice/rehearse	I could have time to jot down the points I wanted to include and words and grammar I needed to use.
6. Improve lexical/syntactical choice	I thought of the words and sentence patterns I could use so I could use more variety of expressions in the essays.
7. Improve accuracy*	It could help me reduce errors.
8. Improve fluency*	1) It helped me write faster; 2) I could make fewer corrections when writing. It helped me write more smoothly.
9. Help recall*	I could have time to think about the points to include in the essays so I would not forget the important points when I wrote.
10. Identify problems*	I could try the words or sentences and if I found that I could not use them correctly I would not use them in the essays.

Note: Categories are adapted from Ortega (2005). An asterisk indicates a new category that emerged from the data.

Table 57 Reported benefits of pre-task planning

Benefits of pre-task planning	Questionnaire results (n=19)		Interview results (n=14)	
	Raw	Percentage	Raw	Percentage
1. Help organize thoughts	10	53%	5	36%
2. Help formulate thoughts	1	5%	2	14%
3. Help with structure*	11	58%	5	36%
4. Improve overall content	4	21%	4	28%
5. Practice/rehearse	4	21%	2	14%
6. Improve lexical/syntactical choice	5	26%	5	36%
7. Improve accuracy*	4	21%	2	14%
8. Improve fluency*	5	26%	11	79%
9. Help recall*	2	10%	3	21%
10. Identify problems*	1	5%	3	21%

Note: Categories are adapted from Ortega (2005). An asterisk indicates a new category that emerged from the data.

The top four most reported benefits of pre-task planning are: *help organize thoughts* (53% in questionnaire and 36% in the interview), *help with structure* (58% in the questionnaire and 36% in the interview), *improve lexical/syntactical choice* (26% in the questionnaire and 36% in the interview), and *improve fluency* (26% in the questionnaire and 79% in the interview). In responding to the post-task questionnaire, six pre-task planners did not report in what ways planning helped them and the scores they gave on the five-point scale ranged from 1 to 3, indicating that they did not find planning very helpful. It should also be noted that participants gave more than one reason for why pre-task planning was helpful. That is why the percentages do not add up to 100%.

In general, participants held positive attitudes toward planning. In the post-task questionnaire (Appendix E Q14), participants were asked to choose if they would like to have planning time prior to writing or start writing right after they read the task instructions. Among the 25 participants in this group, only three (12%) indicated that they would prefer not to have time to prepare before task.

### **7.6.2 Learners' perception on the online planning and no planning conditions**

Online planners' and non-planners' views on their respective task completion conditions were gathered through both the post-task questionnaire and the interview. The question asked "How did you feel about having 15 minutes (for NP)/25 minutes (for OLP) to finish writing without any time to prepare in advance?" Answers to the question mostly centered on the following aspects: 1) whether the time was adequate for them to complete the tasks, 2) benefits for their writing, 3) the perceived limitations of writing without preparation time, and 4) which task completion condition they would prefer if they could choose.

Commenting on the length of time for task completion, most students in OLP (90%) answered that 25 minutes were enough to complete the task. However, one learner pointed out that "it was not hard to complete the tasks but hard to write well." This could be understood in relation to some students' complaints about feeling restricted by their L2 knowledge. For example, another student wrote that "it is hard to write well because I do

not have enough vocabulary and complex sentence patterns to express myself.” In contrast, for most participants in the NP group (66%) 15 minutes were insufficient for them to complete the tasks.

For the perceived benefits of the task conditions, interestingly a few students in both groups mentioned that writing without preparation before the task promoted writing speed (four or 16% in OLP and seven or 28% in NP). An examination of the temporal fluency of these students in the treatment sessions found that this was true for four of the students in the NP group (4/7) but not true for the students in the OLP group. This indicates that there could be a gap between what they felt and what the data showed in terms of their performance. A student in the OLP group further commented that it was a good way to test writing ability. One student in the NP group wrote that “writing under pressure could train us to think quickly and is good for our writing ability in general.”

With regard to the limitations of writing under their respective task completion conditions, two learners (7%) in the OLP group commented that it was not helpful for increasing vocabulary or improving quality of their essays. Half of the participants in the NP group (50%) thought writing under time pressure was not good for developing their writing abilities for various reasons. Some of the most frequently mentioned reasons are: 1) “Time was short and I could not think of grammar and I could not express my ideas fully. So it was not helpful for my writing ability.” 2) “Time was too tight and I did not have time to think, so the language could not be precise.” 3) “Time was short so I could only use words and expressions I am very familiar with. I did not have chance to try low frequency words.” To sum up, they thought the time they had for writing was too short for them to write good essays so it was not helpful for enhancing their writing ability in general.

When asked if they could choose which task completion condition they would prefer, half of the learners in the OLP group chose to have 25 minutes to write while the other half chose to have 10 minutes to prepare and 15 minutes to write before writing. One learner wrote, stating the reasons for his/her choice, “We should have some time to think before starting to write and to organize thoughts. If not, we will easily miss important

points out. After writing down the sentences we may come up with better expressions but did not have a chance to make changes.” As for NP, 69% of participants chose to have some time to prepare before writing and 31% chose ‘start writing right after reading the instructions’.

In short, writing without time to plan before task was generally considered not helpful for developing L2 writing ability.

## 7.7 Summary of results

This section summarizes the main findings from the post-task questionnaire and semi-structured interview. Reports from the pre-task planners show that during planning more students focused on planning *content* and *language* than *structure*. When planning the language to be used in the essays, more students paid attention to lexis than to grammar or syntax. In terms of the planning strategies they employed, the most reported strategies were *previewing*, *organizational planning*, *rehearsing*, *writing as prompt/trial*, *translating*, and *lexical/syntactical search*. In response to the question on how much of the pre-task plan was applied in the essays, the pre-task planners estimated that the amount of plan applied ranged from 50% to 80% in general. An analysis of the planning notes verified the participants’ report. The percentage of planned words used in the essays ranged from 58% to 71%. Moreover, repeated measures ANOVA tests revealed that the total number of words written on the planning sheets and the percentage of planned words applied increased significantly over time. With respect to planning styles, a general trend of changing from Type A to Type B was found, indicating that as they practiced more the participants tended to write more complete sentences on planning sheets as opposed to words and phrases. When commenting on the opportunity to plan prior to task, most of the pre-task planners (88%) considered having time to prepare before writing helpful and the most reported benefits of pre-task planning included *improved fluency*, *improved structure*, *improved lexical choice*, and *help with organizing thoughts*.

Turning to the writing process, because OLP and NP did not have the opportunity to prepare before task, they allocated a little time to engage in *previewing* and

*organizational planning* before they actually started writing. With regard to while-writing focus, almost all participants reported focusing on both *content* and *language*. For PTP and OLP, the reported focus on *language* was greater than focusing on *content*. When focusing on language, more students in PTP reported paying attention to the accuracy aspect of language while more students in OLP reported paying attention to complexity.

In terms of writing strategies, the most reported strategy by all groups was *improvising*. However, more participants in the OLP and NP than in the PTP group relied on this strategy. Two strategies used more by PTP than the other two groups were *lexical/syntactical search* and *elaboration*. OLP used *rereading*, *pausing*, and *editing* more than the other two groups.

## **7.8 Discussion**

In this section, the results from the post-task questionnaire data and the interview data will be discussed following the order the research questions were laid out and in relation to previous studies that investigated the process of planning and the effects of task planning on L2 writing, which have been reported in previous chapters. In section 7.8.1, how successful each planning condition was operationalized is discussed first. This is followed by a discussion of the results for RQ3. In the last section, how social context exerted influence on the process of planning and the effects of planning are discussed.

### **7.8.1 Operationalization of planning**

Since the present study set out to investigate the effects of pre-task planning and online planning on L2 writing, it is necessary to establish the extent to which the two planning conditions were successfully operationalized. In the next few paragraphs the effectiveness of pre-task planning will be commented on first, followed by the effectiveness of online planning.

The pre-task planning was operationalized as providing 10 minutes for the participants to engage in solitary planning before they performed the writing task. The length of planning time had been established and commonly used in previous planning



studies (e.g. Crookes, 1989; Foster & Skehan, 1996; Yuan, 2001; D. Li, 2004; Wang, 2009). In the interview, the pre-task planners of the present study were asked if they used the whole 10-minute planning time (see Appendix F Q17). Thirteen out of fourteen pre-task planners, who were interviewed, reported having sufficient time to plan and 52% of them said that they used the whole 10-minute planning time while the other half reported having used about 8 minutes to plan. This showed that the length of planning time was appropriate for the participants.

To find out what the students were doing during pre-task planning time, a questionnaire and an interview were conducted. In these sessions, students reported that they engaged in various planning activities (e.g. organizing ideas, rehearsing language, etc.) during that period. These are some evidence of pre-task planning. Moreover, during planning time learners were given planning sheets, on which they could make notes. Their planning notes also provided some evidence for what they were doing during the 10 minutes. It was found that on their planning notes they jotted down words, phrases and some complete sentences needed for completing the tasks (see Section 7.4.2). These were also evidence of pre-task planning. Therefore, it can be concluded that the operationalization of the pre-task planning condition appears to have been successful.

With regard to online planning, the present study operationalized it as providing the online planners with 10 more minutes than participants in the other two groups to complete the tasks. This was done in the expectation that learners in this group would use the extra time for online monitoring and editing. However, how to make use of the time was in the hand of the writer and was hard to control. In the post-task questionnaire and interview, the online planners reported using on average 2-3 minutes to plan before they started writing, leaving around 22 minutes to write (see Section 7.5.1). This means they used only 7 more minutes to write than the other two groups. However, evidence from the post-task questionnaire and interview suggests that the online planners did engage in online planning. In the interview, more than half of the students reported employing rereading (54%), pausing (69%), and editing strategies (69%) during task performance, which were indicators of online planning. Moreover, as has been reported in Chapter 5,

the online planning group produced significantly more words in the treatment sessions than the other two groups ( $F=4.819$ ,  $p=.01$ ), suggesting that this group spent the extra time on writing the essays. Thus, the operationalization of the online planning condition could be concluded as relatively successful since they did have some extra time (i.e. approximately seven more minutes) for completing the tasks and the time was spent on writing for the tasks.

### **7.8.2 The process of planning**

The process of planning in the present study was investigated in terms of the focus of planning and the strategies employed while planning. The results showed that during the pre-task planning time, learners' focal attention was on conceptualizing the propositional content, translating ideas into words, and structuring the essays. This finding corroborates R. Ellis and Yuan's study (2004), in which most of the pre-task planners reported having focused on planning the content of the story, working on the language needed to encode the story, and rhetorical planning (e.g. outlining the key events and establishing links between pictures).

With regard to the strategies employed during planning, the most reported strategies in the present study were two metacognitive strategies (*previewing* 57% and *organizational planning* 92%) and five cognitive strategies (*rehearsing* 79%, *writing as prompt/trial* 100%, *lexical/syntactical search* 64%, *translating* 50%, and *elaboration* 43%)<sup>18</sup> (see Table 49). These results of the present study were similar to those of some previous studies. In terms of metacognitive strategies, *organizational planning* was one of the most reported strategies in R. Ellis and Yuan (2004), Ortega (2005), and Guará-Tavares' (2008) studies. In light of the fact that R. Ellis and Yuan's study (2004) and the present study used writing tasks whereas Ortega (2005) and Guará-Tavares' (2008) studies used oral tasks, it seems that most L2 learners would engage in organizational planning when preparing for tasks, no matter if they are oral tasks or writing tasks.

---

<sup>18</sup>The percentages quoted are from the interview data.

With respect to the cognitive strategies, *rehearsing*, *writing to prompt/trial* (this is similar to writing/summarizing/outlining in Ortega's (2005) and Guará-Tavares' (2008) studies), and *lexical/syntactical search*<sup>19</sup> were also the most reported strategies in Ortega (2005), and Guará-Tavares' (2008) studies. The frequent report of the *translating* strategy was also found in Ortega's study (2005) but not in Guará-Tavares' study (2008), whereas the frequent report of the *elaboration* strategy was found in Guará-Tavares' study (2008) but not in Ortega's study (2005). The corroborative results for pre-task planning strategies of the present study with these two studies indicate that ELS learners use similar strategies to prepare for language production, irrespective of modalities.

The planning strategies reported in the present study suggest that pre-task planning lends opportunities for learners to engage in advance planning, retrieval, and rehearsal operations. This fits Hayes and Nash's (1996) description of planning in writing (see Chapter 3) in that advance planning corresponds to process planning and retrieval and rehearsal correspond to text planning in their model.

During pre-task planning time, students made a plan for the task using various planning strategies. How much of their plan was applied in the subsequent writing was also investigated. Participants reported that they were able to apply 50%-80% of their plans (see section 7.4). An analysis on their planning notes verified their report. Moreover, it was found that over time some students gradually changed their planning style from listing words and phrases to writing more complete sentences. Their application of plans increased significantly from Time 1 to Time 4 writing. This suggests that students felt that writing complete sentences could help them write better in the subsequent task. That is, this strategy might enhance students' memory of the linguistic items they had planned, thus helping with application of plans in subsequent writing. The effectiveness of this strategy needs to be further investigated by future studies.

Turning to how these planning strategies influenced task performance, some of the reported strategies (see section 7.3) could have affected the results of planning on writing

---

<sup>19</sup> This was termed *lexical search* in Guará-Tavares's study (2008) and was put in the category of *lexical compensation* in Ortega's study (2005).

performance, which will be discussed below. In the treatment sessions, PTP significantly outperformed NP in lexical sophistication ( $p=.023$ ). This could possibly be attributed to the retrieval and rehearsal operations students engaged in during planning time. As discussed earlier, many students reported having focused on language planning (71%) and half of the interviewees specified that they spent more time and energy retrieving words than sentence patterns needed for subsequent writing. Moreover, *lexical/syntactic search*, *rehearsing*, and *writing as prompt/trial* were among the most reported strategies. In some cases, students ventured to use vocabulary they were not familiar with at risk of sacrificing accuracy. The following excerpt illustrated this: “I would try to use accurate forms within my ability. But if I had to use a certain word I was not familiar with, I would still use it even though I was not sure I used it correctly.” As a result of employment of the above mentioned strategies, PTP benefited from pre-task planning in terms of lexical sophistication. This indicates that pre-task planning could provide an opportunity for learners to stretch their interlanguage to use more challenging words. It supports VanPatten’s contention (1996) that when processing meaning learners attend to lexis before they process grammar.

However, this benefit of pre-task planning did not cause development in lexis over time as seen from the insignificant within-group difference of PTP in the testing sessions. This suggests that the effect of pre-task planning on the use of vocabulary is limited to the immediate performance. For L2 development in lexis, further assistance such as instruction or providing more input might be needed.

The effects of pre-task planning also demonstrated in repair fluency. In the treatment sessions, PTP significantly outperformed both NP ( $p=.003$ ) and OLP ( $p=.023$ ) in this aspect in the treatment sessions. This might be caused by the use of rehearsing strategy and the change in planning style. 79% of the pre-task planners reported having engaged in rehearsing during pre-task planning. In addition, they gradually changed from listing words and phrases on planning sheets to writing more complete sentences over time, which seems to have enhanced their memory of pre-task plans and helped with plan application in the subsequent task. Because of these, they were more confident when

writing the essays, thus resulting in making fewer self-corrections than the other two groups.

In terms of development in this aspect, there was a trend for reduction in dysfluency of PTP in the testing sessions as seen from the mean scores (see section 6.3.4). PTP's reduction in dysfluency caused a significant between-group difference ( $p=.012$ ,  $PTP < NP$ ), indicating that providing an opportunity to rehearse before writing is beneficial for reducing dysfluency over time.

### **7.8.3 Students' perceived benefits of pre-task planning**

The perceived benefits of planning were classified into ten categories (see section 7.6). Four of them were reported most frequently: *help organize thoughts* (53% in the questionnaire and 36% in the interview), *help with structure* (58% in the questionnaire and 36% in the interview), *improve lexical/syntactical choice* (26% in the questionnaire and 36% in the interview), and *improve fluency* (26% in the questionnaire and 79% in the interview). Of the four benefits, three were reported in previous studies. *Help organize thoughts* and *improve lexical choice* were reported in Ortega's study (2005) and *improve fluency* was reported in Wigglesworth and Elder's study (2010). *Help with the structure* was a benefit uniquely reported in the present study. This might be due to the fact that oral tasks were used in these two studies whereas writing tasks were used in the present study. According to the literature on oral and written language, there are some unique features of each mode (e.g. Biber, 1987; Chafe & Tannen, 1987; Tannen, 1985). For example, spoken discourse is considered to be context-bound while written discourse is 'decontextualized' or autonomous (Olson, 1977). When speaking, the listener(s) is often present and thus the speaker has the advantage of asking questions for clarification when necessary, whereas in written communication the reader(s) is usually not present and the writer has to write to an assumed audience. Relatively speaking, oral discourse focuses more on interpersonal involvement whereas written discourse focused more on information (Tannen, 1985). Moreover, cohesion in oral discourse can be realized by nonlinguistic (i.e. prosodic) cues while in written discourse cohesion must be lexicalized

(ibid). For these reasons, structure of essays, which contributes to clarity and comprehensibility of message, may have been emphasized more in writing classes in contrast with oral classes. This might be a possible reason for the reported benefit of planning on structure uniquely found in the present study.

Interestingly, students' belief that planning benefited structure was not verified by their writing performance. The between-group differences in organization were not significant in the treatment or testing sessions. Many participants simply sequenced their arguments in the order the information for possible inclusion in the essays was presented on the task sheets (see section 5.6.4). The discrepancy between this perceived benefit of planning and students' actual performance in terms of organization reflects their lack of knowledge about how a compare/contrast persuasive essay can be effectively organized. As discussed in Chapter 5, this lack of discourse knowledge may have placed a ceiling on the effect of their efforts to organize the essays.

In contrast to *help with structure*, the perceived benefits of planning on *lexical choice* and *fluency* were demonstrated in students' writing performance. These have been explained in the previous section. It needs to be noted, though, the benefit of planning on fluency was only evident in dysfluency. In the interview, students used the term 'fluency' to refer to both temporal fluency and dysfluency. Some students expressed explicitly that "because of the preparation I was able to write more neatly with fewer corrections", while some others used a more vague term 'smooth' ("I was able to write more smoothly.").

#### **7.8.4 The process of writing**

The process of writing was investigated in terms of the focus of attention and strategies used during task performance. It was found that there were some differences among the three groups. Fewer participants in PTP (36%) reported having focused on content while writing than those in NP (100%) and OLP (54%). This provides some evidence to show that the pre-task planners did engage in content conceptualization during planning time. Thus, they were able to focus more on linguistic encoding of ideas

during writing. This finding corroborates R. Ellis and Yuan's study (2004), in which they found pre-task planners focused on rhetorical planning and content planning before a task and attended more to translating ideas into words during the task. During the task, more participants in their online planning and no planning group focused on content than those in the pre-task planning group. It is also consistent with Whalen and Menard's (1995) finding that in L2 writing learners engage largely in linguistic level processing, in contrast with pragmatic and textual level processing.

This finding can be explained by information processing theory, Skehan's Limited Capacity Hypothesis (see Chapter 2) in particular. Because the pre-task planners focused more on content during planning time (content:76%, language:56%), they were able to allocate more attentional resources to planning language during writing (content: 36%, language 93%), which may have a cumulative effect and contributed to the long-term gain of PTP in overall accuracy, which has been discussed in Chapter 6 Section 6.6.1.

In terms of focusing on different aspects of language, there were also differences between groups. Half of the students in PTP (50%) reported that they paid attention to accuracy during production and more than half of the students in OLP (54%) reported having focused more on complexity, while fewer students in NP (24% on accuracy, 29% on complexity) reported having focused on these aspects. Half of the pre-task planners' attention to accuracy, together with the rehearsal operations they had engaged in before the task, may further explain the long-term gain of PTP in accuracy.

With respect to strategies employed while writing, there were differences between groups in the use of metacognitive strategies (i.e. *previewing*, *organizational planning*, *monitoring*). PTP differed from the other two groups in that there was no report of using organizational planning or previewing strategies in this group whereas more than half of the participants in OLP (54% for organizational planning and previewing) and NP (53% for organizational planning and 94% for previewing) employed the two strategies. This was because both OLP and NP allocated a small amount of time to do some pre-task planning before they started to write. The two strategies were thus more associated with pre-task planning. This finding was consistent with Nakakubo's study (2011), in which it

was found that advance planning (e.g. previewing) largely took place during pre-task planning time.

In the reported use of cognitive strategies during task completion, there were also some differences between groups. There were more reports of using *lexical/syntactic search* strategy in the PTP group (50%) than in the OLP (38%) and NP (12%) group. The use of this strategy by half of the pre-task planners during the task, together with the use of this strategy before the task by most of the students in this group (64%), could have led to the better performance of this group than the other groups in terms of lexical sophistication in the treatment sessions (see Chapter 5 Section 5.6.2 for a discussion).

The second difference between groups was that there were fewer reports of the *improvising* strategy in PTP (36%) than OLP (85%) and NP (65%). This could, to a certain extent, prove that PTP engaged in planning before task, during which period they, at least partly, planned the language needed for later production. Thus, there was less necessity for them to improvise during task.

The third difference was that the OLP reported less use of the lexical compensation strategy (32%) than the other two groups (PTP: 76%; NP: 60%). This was different from Nakakubo's study (2011), in which her online planning group reported using this strategy more than the pre-task planning group. In her study, no explanation was offered for this result. In the present study, that the OLP had more time to retrieve vocabulary online than the other two groups thus less dependence on the lexical compensation strategy could be a plausible reason for the result.

The last distinctive difference between groups was that over half of online planners reported using *rereading* (54%), *pausing* (69%), and *editing* (69%) strategies while very few students in the other two groups reported these strategies. These reported strategies were some evidence to show that these online planners did engage in online planning during the task, which might have influenced their writing. The influence was evident in fluency and accuracy (see sections 5.6.3 and 6.6.2). In terms of fluency, online planning significantly decreased the temporal fluency of OLP in the treatment sessions and caused a higher ratio of self-corrections than PTP. It might also have decreased the temporal



fluency of this group in the post-test. However, the negative effect on temporal fluency was not long-lasting. Ten weeks after the posttest, in the delayed posttest OLP's temporal fluency increased to a level comparable to that in the pre-test.

While online planning had a negative effect on fluency, it had some positive effects on accuracy, although this effect did not show immediately in the treatment sessions. The impact was again accumulative and present in the use of S-V agreement in the delayed post-test. Online monitoring and editing may have helped students in the OLP group identify and correct rule-based production (Skehan, 1998) errors. Among the 69% of the online planners who reported having edited in the interview, most of (except one case) them managed to correct errors in tense, verb forms, and S-V agreement, etc. Through error detection and correction, their awareness of these problematic grammatical areas was also raised, which probably led to more attention to these areas in later production. This could explain why OLP's use of S-V agreement in the posttest and the delayed posttest improved. Using the ideas in the Output Hypothesis (Swain, 1985, 1993, 1995) to explain this result, when producing output learners may choose to search information in their existing linguistic resources in response to the gap they have noticed in their knowledge base. In doing so, their existing knowledge could be consolidated. In the case of OLP in the present study, when writing, students noticed that S-V agreement was problematic and they used their own knowledge to solve the problem. This led to the consolidation of their existing knowledge and they became more accurate in the use of S-V in later production. Consistent practicing under the online planning condition may have facilitated the automatization of this partially acquired form (i.e. S-V agreement), leading to a reduction in error rate.

In conclusion, pre-task planning seems to have provided the students with the opportunity to conceptualize the content of the essay, engage in advance planning, and rehearse the language needed for meaning conveyance. When preparing the language for later production, more attention was devoted to lexical retrieval than to syntax or grammar. As a result, during writing more advanced words were attempted and participants were more confident about their writing, which led to the immediate positive

effects on lexical sophistication and dysfluency. Another result of pre-task planning was that more attention could be allocated to formulation while writing, during which process more attention was paid to accuracy than complexity, as reported by participants in the post-task interview. This yielded some longer-term effects on accuracy.

Online planning seemed to have offered more opportunities to monitor and edit the output. As a result, it had an immediate negative influence on fluency of writing but some positive effects on the use of certain grammatical features (i.e. S-V agreement) that need rule-based processing in a longer term.

### **7.8.5 The influence of the social context**

The present study did not set out to investigate how social factors influences L2 writing. However, it was found that the social context of the study exerted noticeable influence on the process of planning and results of the study. The influence of social context has been largely ignored by most previous planning studies (Batstone, 2005) but is an important dimension that can bring more understanding to why and how the study produced such results. This section begins with a brief description of the planning knowledge and skills of participants and an explanation of the necessity of providing guidance or training for planning. Then how the kinds of classroom activity participants of the present study were used to could have played a role in creating the results will be discussed.

#### **7.8.5.1 Chinese students' lack of training on planning**

As explained in Chapter 1, participants of the present study had not been formally instructed or trained in how to plan for writing. When given the opportunity to plan in the first treatment session, they had no knowledge of effective planning. As one of the pre-task planners revealed in the post-task interview “I did not know how to plan so I simply wrote down some words I might need in subsequent writing. Later, when you (the researcher) were collecting our planning sheets I peeked at the notes of my classmates and found that they wrote some sentences. So the second time I tried to write more

sentences on the planning sheet.” This admission of not knowing how to plan was reported by some other pre-task planners as well. Therefore, this group of participants could be considered as not having much knowledge and skills of planning.

However, planning is complex and involves meta-cognitive and particular cognitive skills. Without training and practice, learners might not be able to bring planning to its due effects on L2 production and learning. As has been discussed in Chapter 3 Section 3.2, planning involves goal setting and problem solving, which requires motivation and meta-cognitive skills (Das, et al. 1996). During formation of a plan, learners need to be conscious of their own cognitive processes (meta-cognitive knowledge) and be able to use their cognitive skills strategically in ways that allow them to achieve the goals defined by a given task. For a writing task, this involves identifying audience and requirements of the task, planning the macrostructure of the text and textual coherence of ideas, and making use of the existing L2 knowledge to translate ideas into linguistic forms. Therefore, planning is complex and needs training to achieve its effects (e.g. training might help learners develop ideas and use language which is linguistically rich).

Given the complex process of planning and Chinese learners’ lack of training, the effectiveness of planning could have been limited by participants’ existing ideas about what planning involves. Unless otherwise altered by the teacher in effective and salient ways, participants might conceivably fall back onto well established ways of responding to instructions to write (plan) in their schooldays.

#### **7.8.5.2 The influence of Chinese educational culture**

The concept of educational culture refers to the “taken-for-granted frameworks of expectations, attitudes, values and beliefs about how to teach or learn successfully” (Jin & Cortazzi, 2006). In the present study it refers specifically to participants’ experiences as pupils in secondary schools. In China, secondary school students have limited opportunities to use language in creative ways. This may find its root in the belief that demonstration, modeling, repetition, and memorization of words and grammatical rules are essential in language learning (ibid). In English classes, learners are supposed to

listen and watch carefully demonstrations and make a strong effort to memorize the input.

Moreover, English teaching is under huge influence of the washback effect of public examinations. This idea is going to be explained in two steps: why the washback effect is huge in China and then how it affects the results of the present study. China has the largest population in the world and educational resources are far from enough to meet the needs of students. The competition for getting into a good university is fierce because a degree from a prestigious university is more likely to bring a good job. To get into a good university, students have to achieve high scores in the National Entrance Exam. Therefore, students study and teachers teach for exams. This is why the washback effect is huge in China. With this background, it could be expected that teaching is very much exam oriented. In this pedagogical practice, creativity is not emphasized because it is usually not evaluated in standardized public exams. As has been explained in Chapter 5, in preparing students for the English writing exam teachers train students to follow the guidance provided on exam paper point by point and to do exactly what they are required to do. This is because for the writing exam, students are often required to write a piece around 150 words, in which several main points (usually three points) have to be included. If any main point is missing, scores will be deducted. This training had an impact on the planning process of the participants of the present study and affected the results of the study.

Again as has been described in Chapter 5, for each task in the present study, some information on the objects under comparison was provided on the task sheets. Although participants were encouraged to come up with new ideas, the majority of them just copied information from the task sheets to complete the task. Moreover, though the information provided on the task sheets was arranged in random order, many participants followed the order to compose their essays. They understood the task as requiring them to put together the given ideas and add necessary details in an organizational pattern already provided. For example, the instructions for Task 4 are as follows.

“You have a pen pal from country X. From his/her letter you find that s/he does not have much knowledge of what your country is like and how your life is now. S/he

thinks you are living like how you were 10 years ago. Write a letter to tell him/her how your life is different from 10 years ago. *Try to write as many aspects and details as you can. The following aspects are just for your reference. You can think of other aspects to write about when you are comparing your present life with that 10 years ago.*"

After this eight aspects for possible inclusion were suggested:

- People's income
- Living conditions
- Variety of commodities in the shops
- Price of commodities, e.g. food, clothing, houses, electronic device etc.
- Education
- Access to information
- Environment, e.g. air quality, water, etc.
- People's life span

The following is the planning notes of one participant.

"pen pal,  
from your letters I found that...  
in your imagination our live condition just like....  
So I'd like to tell you some great change of the decade.  
above all, people's income matters most  
on the other aspects, such as living conditions. Most of us moved to the new houses  
Also, there are more variety of.... And.... And our education is also developed,....  
With the advancement of # computers, we can get the information easier and easier."

Obviously, this student decided to use only the suggested points and followed the exact order they were presented. In the interview, she revealed that she did not spend much time generating or organizing ideas because there was no need to do so. Below is an excerpt of the interview.

Researcher (R): What were you thinking most of the time when you were planning?

Student7 (Cao): I was thinking about the words and sentences I could use to express the ideas.

R: Did you spend much time on what to write?

S7: No, because the points are already given.

R: What about the structure of the essay?

S7: I just followed the order. I spent a little time on thinking about how to begin and how to close.

She was not alone in planning in this way. Ten out fourteen pre-task planners reported in the interview having just copied information from the task sheets and followed the order it was presented.

When students were working chronologically down the bullet points, the task almost became a matter of trying to link the words provided by using additional words so that the passage as a whole makes sense. This amounts to a kind of cloze exercise – a highly controlled exercise (an exercise not a task) which is likely to limit the scope for complexity.

As has been demonstrated above, the training students had received influenced the way they approached and planned for the tasks. This, together with the fact that they did not have prior instructions or training in planning, might have affected the results of the study in ways that have been discussed in Chapter 5 and 6. Though the present study did not set out to examine the effects of educational culture on planning, it has been found to be influential. This finding supports the argument for the importance of taking account of the social context in planning studies and TBLT research (Batstone, 2005, 2012; Ortega, 2007). As Coughlan and Duff (1994) pointed out “the tasks in which humans engage exist within a larger, multi-level segment of human activity and that there is a dialectic relationship between humans and the interaction in which they are engaged” (p.174). That is why “a dynamic perspective should be considered when we analyze linguistic data collected through use of a research task. This entails shifting from a focus on linguistic production alone to a more comprehensive consideration of the activities in which such linguistic production occurs” (ibid, p.174).



## CHAPTER 8 CONCLUSION

### 8.1 Overview

This final chapter, firstly, summarizes the key findings of the present study. It then considers the theoretical implications and pedagogical implications of the research findings. The methodological issues that are raised by the present study will also be discussed. These will be followed by a discussion of the limitations of the study. In the last section, directions for future research in this area are suggested.

### 8.2 Summary of main findings

The three research questions the present study addresses are: 1. Does planning have effects on immediate L2 writing performance? 2. Does repeated practice under planning conditions lead to changes in L2 writing over time? 3. What do learners do during planning and writing time? To answer these questions, a pre-test-posttest-delayed posttest between-group quasi-experimental design was adopted. Students' essays produced under different conditions (i.e. a no planning, a pre-task planning, and an online planning condition) in the four treatment sessions were compared to investigate the immediate effects of task planning. Essays written in the pre-test, posttest, and the delayed posttest were analyzed to answer the second research question. All essays were analyzed in terms of accuracy (correct use of comparative forms, correct use of S-V agreement, errors per 100 words, percentage of error-free clauses), complexity (mean length of T-units, number of clauses per T-unit, lexical variety, lexical sophistication), fluency (number of words per minute, dysfluency), and organization (cohesion and coherence rating, automated scores for local cohesion and coherence, automated scores for global cohesion and coherence).

To answer research question 3, all participants were asked to fill in a post-task questionnaire, answering questions regarding their planning process, writing process, and their perceptions of their respective task completion conditions. Half of the participants



were also interviewed to provide more in-depth information on these aspects. The research findings are summarized in the following sections.

### **8.2.1 The immediate effects of planning**

Regarding the first research question (Does planning have effects on immediate L2 writing performance?), it was found that 1) neither pre-task planning nor online planning had immediate positive effects on accuracy of writing; 2) pre-task planning had some positive effects on lexical complexity but no effect on syntactic complexity while online planning did not have any effect on complexity; 3) pre-task planning had no effect on temporal fluency but reduced dysfluency while online planning had a negative effect on temporal fluency and no effect on dysfluency; 4) neither pre-task planning nor online planning had an effect on organization.

With respect to accuracy, the results for pre-task planning corroborate findings from previous studies on planning in L2 writing contexts (e.g. R. Ellis & Yuan, 2004; Ojima, 2006) and the results for online planning were also in agreement with most of the previous studies (except R. Ellis & Yuan, 2004). Given the above findings, it is suggested that planning does not have an immediate effect on accuracy of L2 writing. However, the following factors could have influenced the results for accuracy in the present study.

1. Learners' understanding of task instructions and the way they approached the tasks. As explained in Chapter 4 & 5, for each task, some words and expressions they might need were provided on the task sheets in order to elicit more varieties of comparative forms and suggest possible ideas for inclusion. They were intended to be a resource-dispersing factor (Robinson, 2001, 2005; Robinson & Gilabert, 2007) that could provide support for task completion. Deviating from the researcher's intention, some participants regarded the word list as the list of must-use words. When this happened, the unfamiliar words on the list became sources of errors. This understanding of task instruction made avoidance of using these words difficult. When they risked using the unfamiliar words, errors occurred. Planning opportunities, pre-task planning or online planning, could not help overcome the language difficulties they encountered during

production. Furthermore, when participants focused on incorporating these words in their essays, attention that could have been otherwise paid to other linguistic forms was reduced. This might partly explain why planning did not enhance the overall accuracy of the immediate performance.

2. The limited guidance to direct learners' attention to accuracy (see Chapter 5 Section 5.6.1). Unlike some previous studies (e.g. Mochizuki & Ortega, 2008; Skehan & Foster, 2005) that found increased accuracy as a result of strongly guided planning, only brief instruction was given to the pre-task planning group to direct them to plan both the content and language. There was no specific suggestion on how to plan the content or language. For the online planning group, simply more time was given to fulfill the tasks. No suggestion was given on how to make use of the time. How to use the time and what to attend to were completely in the hands of the students. Planning might have to be assisted with direction of attention to specific aspects of performance to achieve an effect.

3. Limits of L2 knowledge. Data from the post-task questionnaire and interview revealed that students' limited L2 knowledge might have put a ceiling on the effects of planning in terms of overall accuracy. Though many of the online planners (69%) had time to monitor and edit their essays, only some of them (38%) were able to correct some errors and the number of errors they corrected was small (two or three). Some of them (31%) could not identify their own errors, thus unable to correct any error. This shows the limit of their L2 knowledge, which could not be overcome by planning opportunities. In other words, planning could only improve the use of linguistic forms that learners have full declarative knowledge of but have not been fully proceduralized. For errors that are caused by a lack of knowledge, pedagogic measures, such as instruction or corrective feedback, rather than planning might be more helpful.

4. Limitation with measurement. As discussed in Chapter 5, there were problems with both specific measures and general measures for accuracy. The specific measures were found to enjoy a high rate of accuracy at the outset of the study (above 90%), leaving little space for improvement. Gains in the use of the target structures, if any, were hard to achieve statistical significance. The general measures for accuracy (i.e. error/100

and %error-free) used in the present study were based on instances of errors. Types of errors and seriousness of errors were not assessed. Thus, the picture regarding accuracy was not complete.

In terms of complexity, the results for syntactic complexity accord with most of the previous studies on planning in L2 writing (e.g. Al-Humaidi, 2008; D. Li, 2004). It seems that planning does not have much effect on complexity. Nevertheless, the following factors could have influenced the results in the present study.

1. Task complexity. As described in Chapter 4, there were quite a few resource-dispersing variables (e.g. provision of information for possible inclusion in the essays, planning time) in the tasks that reduced task complexity. One such variable that might have exerted influence on complexity in particular was the provision of information on the objects under comparison on the task sheets. This might have led to similarity in content of the essays produced by all three groups of participants, which in turn might be part of the reason for the similar scores in syntactic complexity. A comparison of the points made in the essays and the information provided on the task sheets found that most of the ideas included in the essays (76.7% on average) were copied from the given information on the task sheets, which indicates that participants did not try to be more creative with the content even when planning time was available. In other words, provision of information on the task sheets reduced the task demand on planning the content. Therefore, the potential benefit of pre-task planning and online planning for conceptualization was weakened and the benefit for syntactic complexity could not be seen. This, to a certain extent, supports the argument that the effects of planning would be more evident in cognitively more challenging tasks (e.g. Foster & Skehan, 1996; Kellogg, 1990; Wigglesworth, 1997).

2. Limitation with measurement. The measures for syntactic complexity (i.e. C/T and MLT), though having been used widely in planning studies, have some problems when used for measuring syntactic complexity of written language. Biber et al. (2011) argue that measures for T-units and clause subordination are more characteristic of spoken language than academic written language. Some other studies (e.g. Knoch, 2007; Lu,

2011) found that C/T could not distinguish essays written by learners of different proficiency levels, suggesting that it is not a good index for written language complexity. The problems with measurement might be a reason for the lack of effect on syntactic complexity.

3. Limitation in L2 knowledge. In the interview, some participants in the online planning group expressed their frustration at not being able to use a wider variety of words and sentence patterns because of the limits of their L2 knowledge. When writing, they may have noticed a gap between what they would like to say and the linguistic resource they had available. Planning did not help close this gap. This might partly contributed to the lack of effect on complexity, despite the efforts students in the online planning group made to use more complex language.

4. For the positive effect of pre-task planning on lexical complexity, the opportunity to retrieve vocabulary during the pre-task planning period has been argued to be the reason for it. Lexical search (64%) was one of the most reported planning strategies in the post-task interview and 50% of the pre-task planners reported that retrieving vocabulary was one of their planning foci. Moreover, *improve lexical choice* was one of the most reported benefits of pre-task planning as perceived by the participants. The opportunity to retrieve vocabulary before a task may have given this group the advantage of using more advanced words than NP.

Regarding fluency, the results for pre-task planning on temporal fluency were different from some previous studies (e.g. R. Ellis & Yuan, 2004; M. D. Johnson, 2011) but consistent with three other studies (D. Li, 2004; Ong & Zhang, 2010; Pu, 2009). Difference in task type was hypothesized to be the cause for the difference in results. Most previous studies used narrative tasks while Ong and Zhang's study and the present study adopted argumentative tasks. Studies on narrative and argumentative schema suggest that narrative discourse is relatively easier to be recalled later than argumentative discourse (e.g. Britt et al., 2007). The difficulty in recalling and executing the pre-task plan might influence temporal fluency of writing. This hypothesis needs to be further researched by future studies.

Apart from this hypothesis, the task feature mentioned above (provision of information on object under comparison) has been posited to explain the result. In theory, pre-task planning provides an opportunity to prepare the content and structure of the essay and with a planned content and structure pre-task planners should be able to write faster than non-planners. However, provision of information on objects under comparison reduced the potential advantage of pre-task planning on conceptualization. Thus, the predicted increased production speed in the planning condition was not evidenced.

The result for the effects of pre-task planning on repair fluency corroborates findings of R. Ellis and Yuan's study (2004). In the present study, the opportunity to engage in content planning and language planning has been argued to cause the significant positive effect of pre-task planning on repair fluency (i.e. reduced dysfluency). There is evidence from the questionnaire and interview data to show that during the pre-task planning time, participants planned for both content and language, at least partly, for the subsequent production. This alleviated the online processing load and might have given students more confidence in writing, leading to the significant decrease in dysfluency.

For the negative effect of online planning on temporal fluency, having less time pressure and more opportunities to engage in online monitoring and editing were considered to be the reasons. There is evidence to show that participants in the online planning group felt less pressed for time and engaged in more online monitoring and editing than those in the other two groups (see Chapter 7). These factors might have affected the temporal fluency of the online planners.

Turning to organization, neither pre-task planning nor online planning had immediate effects on organization of the essays. These results were consistent with Kellogg's study (1987) on the effects of pre-writing strategies on L1 writing but not directly comparable with studies on planning in L2 writing since different measures for organization were adopted (see Chapter 5 Section 5.6.4). The following reasons were forwarded to account for the results.

1. The provision of information on task sheets also might have influenced the results for organization. Although the information on objects under comparison was provided in

a random order and was not meant to be a suggested organizational pattern, many participants just followed the order in which the information was given to structure their essays. Moreover, the compare/contrast type of discourse allows for a topical order, making it possible for the students to use the order of given information as the organizational pattern for their essays. This was part of the reason for the insignificant between-group differences in organization of the essays.

2. Limits of L2 knowledge could also explain the lack of effects of planning on organization. As has been explained in Chapter 5, participants of the present study had never been formally instructed on how to write essays in English. Therefore, their L2 discourse knowledge was limited. With limited L2 discourse knowledge, students who had planning time were not likely to produce significantly more cohesive and coherent essays than those who did not have planning time.

The results for the effects of planning on immediate L2 writing performance of the present study, together with results from previous planning studies, suggest that fluency seems to be more susceptible to influence of task planning than the other aspects of language performance (i.e. accuracy, complexity, organization). Complexity is more related to L2 knowledge *representation*, since it is regarded as “the scope of expanding or restructured second language knowledge” (Wolfe-Quintero et al. 1998, p.4 cited in Housen & Kuiken, 2009). Changes in complexity occur in the first stage in Skehan’s (2000) model of interlanguage development (see Chapter 2), which involves extending and restructuring of L2 system. Accuracy and fluency are primarily related to *control* over L2 knowledge (Skehan, 2000) in that improvement in accuracy represents learners’ enhanced ability to use L2 that conforms to the target language norms and gains in fluency reflects increased speed and ease with which learners access existing L2 knowledge in real-time communication. Seen from the results presented previously, it seems that the executive *control* over L2 knowledge as opposed to L2 knowledge *representation* tends to be more easily influenced by task planning.

### 8.2.2 The effects of planning on L2 writing development

The second research question asks ‘Does repeated practice under planning conditions lead to changes in L2 writing over time?’ Results show that 1) both pre-task planning and online planning had some positive effects on accuracy in some measures over time; 2) neither pre-task planning nor online planning had any effect on complexity; 3) both pre-task planning and online planning had negative effects on temporal fluency but the negative effect for online planning did not sustain; and 4) neither pre-task planning nor online planning had any effect on organization of the essays over time.

Result 1) (for accuracy) did not follow from the results for the immediate effects of planning of the present study and many previous studies. Two main reasons below were suggested to account for the result.

1. The space provided by pre-task planning to attend to accuracy might have led to the significant gain in overall accuracy of the pre-task planners. In the post-task interview, a higher percentage of pre-task planners reported having paid attention to accuracy than non-planners and online planners, suggesting that pre-task planning offered more opportunities to attend to accuracy. The sustained attention to accuracy achieved some noticeable effects over time. This supports Ortega’s argument (1999) that the benefits of sustained attention to formal aspect of the language are indirect and accumulative.

2. The significant gain in the accuracy rate of S-V agreement of the online planning group could be attributed to the opportunity to engage in online monitoring and editing. Through repeated error detection and correction during online monitoring and editing, the awareness of the problematic linguistic areas was raised and enhanced. This led to significant increase in accuracy in S-V agreement in the longer-term.

Result 2) (for complexity) fits with the immediate effects of planning on complexity found in the present study and some previous studies (e.g. Al-Humaidi, 2008; M. D. Johnson, 2011). The explanations for the result are outlined below.

1. As has been argued in Chapter 5, the provision of ideas on the task sheets might have weakened the immediate advantage of planning for syntactic complexity. It

discouraged creativity in writing for the tasks, leading to similarity in content of participants' written texts, which in turn affected language complexity. This might have limited the need and chance for the participants to use more complex language. It is not surprising that repeated practice under planning conditions on writing tasks with such a feature did not lead to significant gains in this aspect over time.

2. Limitation with measurement discussed in Chapter 5 and 6 could also be part of the reason for the lack of gains in complexity. That is, the two measures used for syntactic complexity might not be discriminating for assessing L2 written language.

Result 3) (for fluency) for pre-task planning did not follow from the immediate effects of planning in the present study and previous studies (e.g. R. Ellis & Yuan, 2004; M. D. Johnson, 2011) while the result for online planning was consistent with these findings. The results were explained as follows.

Some habits may have been developed through repeated practice under the pre-task planning and online planning conditions in the treatment sessions. When they were carried over to the testing sessions, negative effects on temporal fluency were evidenced. For the pre-task planning group, a habit of rehearsing mentally before actually starting to write for the task might have been developed through repeated practice of writing under the pre-task planning condition. For the online planning group, a habit of monitoring and editing might have been formed. These habits have been argued to be the causes for the significant decrease in temporal fluency of the pre-task planning and online planning groups.

Result 4) (for organization) fits with finding for the immediate effects of planning of the present study. The reasons offered to account for the immediate effects of planning (see Chapter 5) were applicable to the longer-term effects of planning. That is, provision of information on the objects for comparison became a suggested organizational pattern for many participants, which reduced the necessity to work on the structure of essays. Students' limited discourse knowledge could not be overcome by the planning opportunities. These two might be the reasons for the lack of effect on organization over time.



### 8.2.3 What learners do during planning and writing

The third research question of the present study was “What do learners do during planning and writing time?” The question was investigated in terms of 1) the focus of planning, 2) strategies students employed while planning, 3) application of their plan, 4) their planning styles, 5) strategies students employed during writing, and 6) their perceptions of the task completion conditions.

Post-task questionnaire and interview data reveals that 1) pre-task planners focused mostly on conceptualizing content, translating ideas into language, and planning the structure of the essay during pre-task planning time; they paid more attention to vocabulary than sentence patterns or grammar during language planning. This finding was consistent with Ellis and Yuan’s study (2004) and indicated that content, language, and structure were three aspects learners normally prepare for the subsequent task.

2) The most reported planning strategies were *previewing*, *organizational planning*, *rehearsing*, *writing as prompt/trial*, *lexical/syntactic search*, and *translating*. These strategies were also frequently reported in other studies (e.g. Ortega, 2005; Guar-Tavares, 2008), suggesting that L2 learners employ similar strategies when preparing for L2 production. Furthermore, pre-task planning provides opportunities for advance planning and retrieval and rehearsal operations.

3) Pre-task planners reported being able to apply 50%-80% of their plans in subsequent writing and this was verified by an analysis of their planning notes and essays. Moreover, it was found that the percentage of applied plan increased from Time 1 to Time 4 writing.

4) The planning style of some pre-task planners gradually changed from jotting down mostly words and phrases on the planning sheets to writing more complete sentences from Time 1 to Time 4 writing as well. This might contribute to the increase of plan application.

5) With respect to the strategies students employed while writing for the tasks, some differences between groups were found. The first was that fewer pre-task planners

reported having focus on content than online planners and no-planners. Another was that most of the online planners and no-planners reported having engaged in *advance planning*, while the pre-task planners did not report doing so during writing. The third difference was that more students in the OLP group and the NP group employed the *improvising* strategy and less *elaboration* strategy than those in the PTP group. These differences could be attributed to the fact that PTP had time to prepare before task while OLP and NP did not have the time. This partly proved that operationalization of the pre-task planning condition was successful.

The fourth difference was that half of the pre-task planners reported having paid more attention to accuracy than complexity while half of the online planners specified that they paid more attention to complexity and very few no-planners reported having attended to either accuracy or complexity. Moreover, there were more reports of lexical/syntactical search in the PTP than in the OLP and NP group. The use of these strategies might have contributed to the long-term gains in overall accuracy (see Chapter 6) and the better immediate performance in lexical sophistication of PTP (see Chapter 5).

The last distinctive difference was that more students in the OLP group used *rereading, pausing, and editing* strategies than those in the PTP and NP group. These strategies were indication of online planning, thus proving the relative success of the operationalization of the online planning condition. They were also associated with the significant gains in the use of S-V of OLP over time.

6) The most reported benefits of planning were *improve fluency, help structure, help organize thoughts, and improve lexical choice*. The perceived benefits for fluency and use of lexis were consistent with results for the effects of planning (see Chapter 5). However, the perceived benefit for the structure of the essays was different from the results. This indicates a gap between their perception and their actual performance.

### **8.3 Theoretical implications**

Results of the present study have implications for some theoretical issues, which will be discussed in the following sections. Some of these theoretical issues overlap in part.

### 8.3.1 L2 skill development

One major finding of this study is that repeated practice under planning conditions has some, though limited, positive effects on accuracy of L2 written production. This has implications for the claims of Skill Acquisition Theory, which predicts that L2 development follows a path of changing from declarative knowledge to procedural knowledge and a slow automatization of that knowledge. Practice plays a central role in the whole process. Evidence of development would be reduction in reaction time and error rate. As for what kind of practice would be helpful, it is contended that practice should aim at developing automatic processes (DeKeyser, 1998) and it should engage learners in comprehending or expressing real thoughts. Task-based research has attempted to address the issue of how to involve learners in communicating real thoughts by using tasks. It has been found that different tasks and task conditions (e.g. having planning vs. without planning) affect accuracy, complexity, and fluency of L2 performance to different degrees. However, this research has only looked at the effects of different task conditions on immediate performance. How repeated practice under different task conditions affects L2 development process has been overlooked. This is the major issue the present study looked into. The design of the study allows for an examination of L2 writing development as a result of different practice conditions.

To establish different practice conditions, the present study created three task conditions: pre-task planning, online planning, and no planning. Results show that the pre-task planning condition induced 'focus on form', which refers to learners' selective attention to linguistic forms while attempting to communicate (R. Ellis, 2005b). The post-task questionnaire and interview data reveals that pre-task planners did attend to linguistic forms during planning, in particular half of the participants paid attention to accuracy during writing time. The online planning condition, which allowed more time for students to monitor and edit their output raised their awareness of the problematic linguistic forms. There is evidence from the questionnaire and interview data to show that OLP engaged in more online monitoring and editing than PTP and NP.

The results of repeated practice under these task conditions are that PTP's overall accuracy and OLP's accuracy of S-V increased over time. This points to the conclusion that practicing under conditions that provide more opportunities to focus on form (attend to accuracy in the case of pre-task planning) and increase learners' awareness of specific problematic linguistic forms (the case of online planning) might be more conducive to develop automaticity than a condition that provides less such opportunities.

However, conditions that induced learner-driven focus on form (planning conditions) might not produce immediate effects, as shown by the effects of planning on immediate performance in the present study. This supports Ortega's (1999) argument that the benefits of focusing on form during pre-task planning might not receive an immediate payoff with respect to accuracy. This also confirms the prediction in Skill Acquisition Theory that automatization is a gradual and slow process.

### **8.3.2 Attention**

The contention that humans' attention is a limited resource is confirmed by the present study. Allocation of attentional resources was found to have more influence on language production when real-time communication pressure is greater. According to information processing theory, humans have limited processing capacity and are unable to attend to all input they are exposed to. In language production, we are not able to attend to all aspects of a task simultaneously. Attention to one aspect may cause the other aspects to suffer.

The results for the process of planning and writing (Chapter 7) of the present study provide some evidence to show that planning can influence learners' allocation of attention during production. For example, pre-task planning allowed learners to engage in retrieval and rehearsal operations during planning time, during which ideas are generated and organized and part of the language needed for meaning conveyance is prepared. As a result, they could allocate more attention to translating ideas into language during production. However, the effects of more attention to language on immediate writing performance were limited to increased lexical sophistication and repair fluency only.

Compared with research that was conducted in oral contexts, the effects of task planning on immediate L2 writing performance, as seen in the present study, seem to be weaker. This indicates that planning, which affects allocation of attention, would have more effects in situations where communication pressure is more acute (i.e. oral production). Its effects will be less evident when there is less communication pressure (i.e. written production). However, prolonged practice producing written language with attention to form would help automatization of L2 knowledge.

### **8.3.3 The Output Hypothesis**

The present study confirms a claim of the Output Hypothesis in part. According to the Output Hypothesis (Swain, 1985, 1993, 1995), producing output could help learners notice a gap between what they need or want to say and what they can say. When a gap is noticed, learners might search their existing L2 repertoire to “close the gap by generating new knowledge or consolidating existing knowledge” (Swain, 1993, p.159) or pay attention to input. In the present study, producing output under planning conditions has been found to facilitate consolidation of existing L2 knowledge but not helpful for extending it. Some participants in the online planning group closed the gap by consolidating their existing knowledge about S-V agreement. With the opportunity to monitor and edit their output, some students in this group noticed their problems with this grammatical feature (see Chapter 7 Section 7.5.3) and reprocessed the relevant knowledge in their L2 system, leading to a significant increase in accuracy of this form over time. This suggests that the condition of producing output plays an important role in L2 development. A less time pressured condition (i.e. with online planning opportunities) might be favorable for developing accuracy of certain linguistic forms.

However, when the linguistic problems are beyond the scope of existing L2 knowledge, planning before or during output might not help to close the gap. In the case of the present study, some participants expressed their frustration at not being able to overcome their limits of L2 knowledge, even when planning time was available. This limit seems to have set a ceiling on the effects of planning. Therefore, to better promote

IL development, planning should be combined with other facilitative methods, such as providing corrective feedback. There is some evidence to show that when learners were exposed to L2 recasts under pre-task planning and online planning conditions, more uptake of recasts and better learning took place (Romanova, 2010).

#### **8.3.4 Task complexity**

Though the present study did not set out to test the Cognition Hypothesis, the results confirmed a claim in the hypothesis. According to the Cognition Hypothesis (Robinson, 2003, 2005, 2007, 2011), increasing task complexity along the resourcing-directing dimension can push learners to go beyond their existing L2 repertoire, extending it to meet the demands of the task, thus engaging the acquisition process. In contrast, resource-dispersing variables, such as planning time, stimulate access to existing L2 knowledge. Increasing these variables alone will decrease task complexity, leading to enhanced performance, but will not cause extension of L2 knowledge base. The two dimensions have to be combined to promote L2 development.

The tasks of the present study only had two resource-directing variables (i.e. required causal reasoning and use of all first, second and third person perspectives) that placed relatively high cognitive demands but four resource-dispersing variables (e.g. provision of planning time and background information) that decreased task complexity. The results of doing these tasks were 1) increased control over the existing L2 knowledge, demonstrated by better immediate performance in repair fluency and longer-term gain in overall accuracy of PTP, and 2) no evidence of L2 knowledge base extension, indicated by use of more complex language, for any of the planning groups. These results confirmed the claim of the Cognition Hypothesis.

The present study also shows that variables that are designed to increase or decrease task complexity could work in an opposite direction affected by learner factors. As discussed earlier, word lists provided on task sheets, which were supposed to alleviate some task demand, somewhat added linguistic demands when understood as must-use words. This leads to the next implication discussed below.

## **8.4 Methodological issues raised by the study**

### **8.4.1 Difficulty of operationalization of planning**

This study has highlighted the difficulty of operationalizing the planning conditions. As has been discussed in Chapter 3, it is impossible to create a situation in which pre-task planning or online planning is precluded. This is because planning is a mental activity beyond the control of the researcher and it is a strategy writers naturally employ when fulfilling a written task. Even under great time pressure, such as the case of the no planning condition in the present study, learners might engage in some pre-task planning before they started to write and some online planning while they were writing. Students' reports in the post-task interview proved this. For example, most non-planners reported having spent one or two minutes and many online planners spent an average of two or three minutes doing advance planning (a pre-task planning strategy) before starting to write. A few participants in the no planning and the pre-task planning group reported having employed the editing strategy (an online planning strategy) while writing. The best the researcher could do was to establish a condition where there were relatively more opportunities for pre-task planning, a condition where there were relatively more opportunities for online planning, and a condition where there was limited space for pre-task planning and online planning. This needs to be taken into consideration when interpreting the results of the present study.

### **8.4.2 Difficulty of measuring L2 writing performance**

The notions of complexity, accuracy, and fluency (CAF) have been employed in SLA research as dimensions for describing L2 performance and indicating IL development. There has been consensus on the usefulness and validity of the three constructs. However, disagreement exists on the operationalization of them and on the best measures for the three dimensions. For example, Wolfe-Quintero et al. (1998) suggested that the best measures for syntactic complexity were clauses per T-unit, number of dependent clauses per total clauses or per T-unit. More recent studies (e.g. Biber et al. 2011; Lu, 2011),

however, show that they might not be good indices for syntactic complexity of written language. Operationalizing CAF first needs clear definitions of the three constructs. Each construct may have sub-constructs. For example, R. Ellis and Barkhuizen (2005) list eight aspects of complexity, including, lexical, interactional, propositional, and various types of grammatical complexity. Even syntactic complexity alone is a multidimensional construct that has several sub-aspects to some scholars (e.g. Norris & Ortega, 2009). Then reliability and validity of all the possible measures and how sensitive they are to changes should be established. Therefore, finding good CAF measures is a huge, complex and demanding job, which is not the focus of the present study.

The present study adopts the most widely used measures for CAF in planning studies to allow for comparison with previous studies. However, it is well acknowledged that the measures have limitations. As discussed in Chapter 5, the measures for accuracy only capture changes in the number of errors, leaving changes in types of errors or seriousness of errors undetected. The measures for syntactic measures have been found to be inappropriate for measuring written language (Biber et al., 2011) and undiscriminating for different proficiency levels (e.g. Knoch, 2007; Lu, 2011). Therefore, the results of the present study should be understood as restricted to the specifically measured aspects of accuracy and complexity.

Apart from the linguistic level analysis (i.e. CAF), the discourse level analysis of learners' essays was conducted in the present study. This also proved to be difficult. For one thing, there is still a lot of disagreement on the definitions of cohesion and coherence. As has been discussed in Chapter 4, there is little consensus on an overall definition of coherence (Grabe & Kaplan, 1996). Some scholars define it as "the relationships that link the ideas in a text to create meaning" (Lee, 2002, p32) while some others argue that coherence does not entirely reside in the text. It is a result of interaction of cohesion and the reader (e.g. Yule, 1996). Given the disagreement in definition and the abstractness and fuzziness in nature of coherence, it is very difficult to assess. Research on L2 writing has largely relied on subjective rating using either holistic or analytical scales. This way of assessing cohesion and coherence is subjective and reliability could be a problem. It is



easy to understand that though following a rubric when rating, each rater may have different judgment on how cohesive and coherent a text is. Recognizing this weakness of subjective rating, the present study used an online program (Coh-Metrix) to assess cohesion and coherence of texts in addition to the subjective rating. Though this program is powerful for analyzing texts and has been used in many published studies (e.g. Baba & Nitta, 2012; Kormos, 2011; Liang, 2006), it still has weaknesses in assessing coherence of text, in particular when coherence is viewed as linked with the reader.

#### **8.4.3 Difficulty of conducting semi-longitudinal quasi-experimental research**

The present study highlighted the difficulty of conducting semi-longitudinal quasi-experimental research. The term semi-longitudinal is used here to refer to the present study, which is shorter than a real longitudinal study but longer than studies that only last for two or three weeks or even shorter. Research of this nature usually has several waves of data collection and requires that participants attend all data collection sessions in order for their data to be included for analysis. When a study spans for a relatively long time, the possibility for students' absence from data collection sessions for various reasons would become high. This would cause a reduction in sample size. It was because of students' absence from data collection sessions that the number of participants whose data could be entered for analysis was reduced from 107 to 81. After screening for proficiency control, eventually 75 participants' data was analyzed and reported.

#### **8.4.4 Use of focused tasks**

To elicit the target structures, focused tasks were used in the present study. The tasks, on the one hand, proved to be relatively successful in eliciting the target structures in that the mean occurrences of comparative forms was over five and the mean occurrences of S-V agreement was around seventeen in each treatment session (see Chapter 5 Section 5.4.1). On the other hand, they had put a constraint on the range of sentence patterns, lexis and text structures that were needed for task completion. This effect was enhanced by the fact that information for possible inclusion in the essays and a list of useful words

and expressions were provided (see Chapter 4). These factors would have some influence on syntactic complexity and lexical complexity of the texts, which need to be taken into consideration in future research.

#### **8.4.5 Influence of sociocultural factors**

The important role of sociocultural factors in SLA research is observed in the present study. Although this study did not investigate how mental activities under different conditions affect language performance and learning from a sociocultural perspective, the influence of the educational culture of participants on study results was found. Results indicated that the social context in which participants of the present study received English instruction played an important role in shaping the way they understood task instructions and the way they went about completing the tasks (see Chapter 7). As has been explained in Chapter 5, many participants made efforts to use the words on the list provided for each task, which led to some errors. This is because through training in writing for exams they have been conditioned to think that failure to include the words on the task sheets would cause deductions from their scores. Without knowledge of this background, it might be difficult to understand why they made such efforts to use these words at the risk of sacrificing accuracy.

This finding supports the argument that “higher forms of mental functioning, including voluntary attention and memory, planning, logical thinking and learning, arise as a consequence of the appropriation of culture” (Lantolf & Pavlenko, 2001, p. 144). Therefore, when analyzing linguistic data contextual factors should be taken into account. Studies that adopt a socio-cognitive approach (Batstone, 2005) might shed more light on our understanding of the complex nature of second language learning.

#### **8.5 Pedagogical implications**

The results of the present study suggest that production practice with planning opportunities is beneficial for developing learners’ automatic language use in the long run in terms of accuracy. This may give teachers (e.g. the researcher) some confidence when

answering students' question---“Can my English be really improved through practicing prepared speech and writing?” Creating task conditions that allow for space to focus on form, such as the pre-task planning condition, will lead to increased accuracy over time. However, teachers should not expect immediate improvement in accuracy as the effect of attention to linguistic form might not be achieved in a short time, since automatization is a slow and gradual process.

The present study also shows that providing a writing condition with opportunities to monitor and edit output may raise learners' awareness of their problematic linguistic areas so more attention may be paid to these areas in later production. This might help further proceduralize the forms over time.

However, teachers need to be aware that different tasks and task conditions may favor or disfavor development of different aspects of task performance (e.g. accuracy, complexity, fluency). For example, writing under planning conditions may disfavor the development of temporal fluency, as shown in the present study. Therefore, teachers should assign tasks to be completed in different conditions to achieve a balanced growth in all three aspects. This could be done by setting focused objectives to be achieved in a teaching period. When having learners practice using less frequent words is the objective, it is favorable to provide planning opportunities for production tasks. However, if temporal fluency is to be promoted, it would be advisable to provide a writing condition with time pressure, such as speed writing, not allowing any planning time.

Teachers also need to be aware that certain task features may promote or limit development in certain area(s). The present study shows that when there is too much support for the task with respect to ideas for possible inclusion, as is the case of the present study, learners tend to be less creative with content, which would restrict the range of language they use. If it were the other way round, there would be more chances for learners to push their interlanguage to limits to use more cutting edge language. Therefore, it is advised that tasks that would encourage creativity in terms of content and language should be used to induce the use of more complex language. This would be more helpful for extending their existing L2 knowledge.

## 8.6 Limitations of the present study

A number of weaknesses of the present study have been identified and will be discussed in the following section.

1. **Sampling:** The present study adopted convenience sampling instead of random sampling, which created a difficulty with proficiency control and affected the generalizability of the results. Three intact classes were used as the three groups in the present study. Though the placement test and the pretest showed that the three classes were not significantly different in terms of L2 proficiency, there were some differences seen from the mean placement test scores of the classes (see Chapter 4 Section 4.3.1) and there were significant differences in two indices of writing performance (i.e. lexical variety and organization rating). In addition, the participants were a group of Chinese first-year students at a university in Beijing. This fact may also limit the generalizability of the findings of this study to other ESL learners.
2. **No control group:** There was no control group that only completed tasks in the testing sessions. If there had been one, the effects of planning and the effects of writing practice could have been isolated.
3. **Length of the testing sessions:** The fact that each testing session lasted twenty five minutes, in which all participants needed to complete an essay without any pre-task planning time, was consistent with the task condition of the online planning group in the treatment sessions. This might have some influence on the results of the posttests.
4. **Performance measures:** As has been discussed earlier in this chapter and in Chapter 5, the measures for accuracy might not provide a complete picture of accuracy and the measures for syntactic complexity might not be discriminating enough for assessing written language. Other measures could be tried in future studies.
5. **Tasks used in the present study:** As has been explained earlier, the use of focused tasks can on the one hand elicit the target structures but on the other hand put a constraint on the range of language learners used. Moreover, the task supporting elements also had affected the results of the study, in terms of accuracy and complexity in particular. It is

acknowledged that the present study used compare/contrast persuasive type of written tasks. The results of this study may not be generalizable to other type of tasks.

6. Interrater reliability for rating cohesion and coherence of the essays: As has been discussed in Chapter 4, the interrater reliability of the two raters hired to rate cohesion and coherence of the essays was not high ( $r=.69$ ). A remedy procedure had to be used to fix this problem (see Chapter 4).

7. Time the retrospective data was collected: The post-task questionnaire was completed a week after the fourth treatment session and the post-task interview was conducted in the following week. Though the time the questionnaire survey and the interview were conducted at the earliest possible time, it was not the best time. It has been suggested that the interval between the time retrospective data is collected and the task be kept as short as possible (Dörnyei, 2007). If the questionnaire survey and the interview had been administered sooner after the treatment was over, more detailed information might be recalled and reported.

8. Data on the process of writing for the posttests: No data on how students went about doing the tests was collected. Therefore, there was no evidence on the writing process they engaged in when writing for the tests. The argument that they carried over their planning and writing habit from the treatment sessions was just a speculation.

9. Use of one type of tasks: Given that only one type of tasks was employed, there was no information available as to whether any effect observed would transfer to other types of tasks.

10. Possible confounding effect of post-task questionnaire and interview: The fact that the post-task questionnaire and interview were administered right after the post-test had a potential confounding effect on delayed post-test results. The results of the delayed post-test should be interpreted with this factor in mind.

11. Limitations of offline protocols: Information on how what learners did during planning and writing time was gathered through questionnaire and interview. This offline measure, while having the advantage of not affecting learners' cognitive processes, has an inherent limitation. That is, it cannot tap moment-to-moment thought processes and some

data might be lost because of recall failure.

## **8.7 Suggestions for future research**

The present study is the first that attempts to find the longer term effects of task planning on L2 writing development. Obviously more studies that gauge the long term effects of planning are needed. Longitudinal studies would shed more light on the role of planning in IL development.

Studies that relate the mental operations of learners during pre-task planning and the subsequent production processes to the subsequent task performance are needed. Correlational studies that involve a larger sample size might be able to provide more insights into how specific planning strategies and writing strategies affect task performance. Information on this might provide a ground for designing training sessions for planning to help learners better use the planning time.

Some studies have investigated how learner difference variables influence the effects of task planning (Guará-Tavares, 2008; Tajima, 2003). Additional studies are needed to detect how individual differences affect planning behaviors and task performance. For example, learners' beliefs about what makes a good essay might affect their choice of what to prioritize in planning, which will subsequently affect their writing performance.

Studies on how planning affects writing performance are still relatively few. More studies are necessary before conclusions on the effects of planning on L2 writing can be drawn. Moreover, CAF has been the major measures for L2 task performance in planning studies. It is time to take into consideration the possible changes in discourse as a result of planning when L2 writing is examined. This is because discourse features are very important, if not more important than linguistic features, for written texts. Therefore, future planning studies should also consider measuring the discourse dimension of task performance.

To date, all planning studies but one (Ojima, 2006) adopted experimental design. More qualitative studies that allow for in-depth discovery of individual differences (e.g. learner belief, motivation, attitude, working memory), planning behaviors, subtle changes

in linguistic aspects as well as discourse aspect of production, and the interaction between these variables would lead to a deeper understanding of whether, how, and why task planning might achieve any effect on L2 production and the role it plays in IL development.

Up to now, planning has not been studied from a sociocultural perspective. Such studies may provide interesting information on how learners' mental functioning is affected by the social contexts and this will further our understanding of the nature of second language learning.

## LIST OF REFERENCES

- Al-Humaidi, M. (2008). *Planning in writing of Saudi EFL learners at the college of languages and translation at KSU*. MA thesis. King Saud University.
- Anderson, J. R. (1978). *Language, memory, and thought*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Anderson, J. R., Greeno, J. G., & Kline, P. J. (1981). Acquisition of problem-solving skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 191-230). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic Press.
- Baba, K., & Nitta, R. (2012). Dynamic effects of repeating a timed writing task in two EFL university courses: Multi-element text analysis with Coh-Metrix. In P. M. McCarthy, & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution*. Hershey, PA: IGI Global.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4, 417-423.
- Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- Baddley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11, 17-34.
- Batstone, R. (2005). Planning as discourse activity: A sociocognitive view. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 277-295). Amsterdam: John Benjamins Publishing Company.
- Batstone, R. (2012). Language form, task-based language teaching, and the classroom context. *ELT Journal*, in press.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.
- Bialystok, E., & Smith, S. M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second language acquisition. *Applied Linguistics*, 6, 101-117.
- Biber, D. (1987). *Textual relations in speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35.



- Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to 'remember' than things that were. *Psychologica Science*, 9, 484-489.
- Britt, M. A., Kurby, C. A., Dandotkar, S., & Wolfe, C. R. (2007). I agree with what? Memory for simple argument claims. *Discourse Processes*, 45(1), 52-84.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Brown, J. S., McDonald, J. L., Brown, T. L., & Carr, T. H. (1988). Adapting to processing demands in discourse production: The case of handwriting. *Experimental Psychology: Human Perception and Performance*, 14, 45-59.
- Butterworth, B. (1980). Some constraints on models of language production. In B. Butterworth (Ed.), *Language Production (Vol. 1)* (pp. 423-459). London: Academic Press.
- Bygate, M. (1996). Effects of task repetition: appraising the developing language of learners. In J. Willis, & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-146). Bath: Heinemann.
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23-48). Harlow: Pearson Education.
- Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37-74). Amsterdam: John Benjamins Publishing Company.
- Byrne, C. (1979). Tutor-marked assignments at the Open University: A question of reliability. *Teaching at a Distance*, 15, 34-43.
- Candlin, C. N. (1987). Toward task-based language learning. In C. N. Candlin, & D. F. Murphy (Eds.), *Language learning tasks* (pp. 5-22). Englewood Cliffs, NJ: Prentice Hall International.
- Carr, T., & Curren, T. (1994). Cognitive factors in learning about structured sequences: Applications to syntax. *Studies in Second Language Acquisition*, 16, 205-230.
- Chafe, W., & Tannen, D. (1987). The relation between written and spoken language. *Annual Review of Anthropology*, 16, 383-407.
- Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition*, 10, 161-174.
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, 18(1), 80-98.
- Chiang, S. (1999). Assessing grammatical and textual features in L2 writing samples: The case of French as a foreign language. *The Modern Language Journal*, 83, 219-232.
- Cohen, J. (1992). A power primer. *Psychologica Bulletin*, 112, 155-159.
- Connor, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English*, 24(1), 67-87.
- Coughlan, P., & Duff, P. (1994). Same task, different activity: Analysis of second language acquisition task from an activity theory perspective. In J. P. Lantolf, & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 173-194).

- Norwood, NJ: Ablex Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 97-185.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Language and Verbal Behavior* 11, 671-684.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367-383.
- Crossley, S. A., Weston, J. L., Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28(3), 282-311.
- Das, J. P., Kar, B. C., & Parrilla, R. K. (1996). *Cognitive planning: The psychological basis of intelligent behavior*. New Delhi: Sage.
- De Bot, K. (1992). A bilingual production model: Levelt's 'Speaking Model' adapted. *Applied Linguistics*, 13, 1-24.
- DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. J. Doughty, & J. Williams (Eds.), *Focus on form in second language acquisition* (pp. 42-63). New York: Cambridge University Press.
- DeKeyser, R. M. (2007a). Conclusion: The future of practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 287-304). Cambridge: Cambridge University Press.
- DeKeyser, R. M. (2007b). Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in a second language* (pp. 1-18). Cambridge: Cambridge University Press.
- DeKeyser, R. M. (2007c). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 219-238). Amsterdam: John Benjamins Publishing company.
- Ellis, N., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes*, 13, 307-336.
- Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, 9, 1-20.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2005a). 'Measuring implicit and explicit knowledge of a second

- language: A psychometric study. *Studies in Second Language Acquisition*, 27(2), 147-172.
- Ellis, R. (2005b). Planning and task-based performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3-34). Amsterdam: John Benjamins Publishing Company.
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30(4), 474-509.
- Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford: Oxford University Press.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59-84.
- Ericsson, A. K. (2002). Towards a procedure for eliciting verbal expressions of non-verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology*, 16, 981-987.
- Ericsson, A. K., & Simon, H. A. (1993). *Protocol analysis: Verbal report as data*. Cambridge, MASS: MIT Press.
- Farahani, A. A. K., & Meraji, S. R. (2011). Cognitive task complexity and L2 narrative writing performance. *Journal of Language Teaching and Research*, 2(2), 445-456.
- Field, A. (2005). *Discovering statistics using SPSS* (Second ed.). London: Sage Publications.
- Flower, L. S., & Hayes, J. R. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 31-50). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Foster, P. (1996). Doing the task better: How planning time influences students' performance. In J. Willis, & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 126-135). Bath: Heinemann.
- Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition*, 18(3), 299-324.
- Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, 3(3), 215.
- Gaies, S. J. (1980). T-Unit analysis in second language research: Applications, problems and limitations. *TESOL Quarterly* 14(1), 53-60.
- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.
- Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernandez-Garcia, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49(4), 549-581.
- George, H. V. (1972). *Common Errors in Language Learning: Insights from English*. Rowley, MA: Newbury House Publishers.
- Gilbert, R. (2007). The simultaneous manipulation of task complexity along planning time and +/- Here-and-Now: effects on L2 oral production. In M. Garcia-Mayo

- (Ed.), *Investigating tasks in formal language learning*. (pp. 44-68): Multilingual Matters.
- Grabe, W. (2001). Toward a theory of second language writing. In T. Silva, & P. K. Matsuda (Eds.), *On second language writing* (pp. 39-58). Mahwah: Lawrence Erlbaum Associates, Publishers.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing*. New York: Longman.
- Grabowski, J. (1996). Writing and speaking: Common grounds and differences toward a regulation theory of written language production. In C. M. Levy, & S. Ransdell (Eds.), *The Science of Writing* (pp. 73-92). Mahwah, NJ: Lawrence Erlbaum Associates.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers* 36, 193-202.
- Guará-Tavares, M. D. G. (2008). *Pre-task planning, working memory capacity, and L2 speech performance*. PhD dissertation. Universidade Federal de Santa Catarina
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing* (pp. 29-56). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of the writing process. In L. Gregg, & E. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, NJ: Lawrence Erlbaum.
- Hayes, J. R., & Nash, J. G. (1996). On the nature of planning in writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing* (pp. 29-56). Mahwah, NJ: Lawrence Erlbaum Associates.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing Company.
- Howatt, A. (1984). *A history of English language teaching*. Oxford: Oxford University Press.
- Hsieh, H., & Shannon, S. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Hulstijn, J. H. (1990). A comparison between the information-processing and the analysis/control approaches to language learning. *Applied Linguistics*, 11(1), 30-45.
- Hulstijn, J. H., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning*, 34, 23-43.
- Hultgren, A. (2009). Using Coh-Metrix to compare cohesion measures between the United States Senators John McCain and Barack Obama. Retrieved from <http://www.essays.se/essay/2af93a3572/>
- Hunt, K. (1966). Recent measures in syntactic development. *Elementary English*, 43,

732-739.

- Izumi, S. (2000). *Promoting noticing in SLA: An empirical study of the output and input enhancement on ESL relativization*. PhD Thesis. Georgetown University.
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the Output Hypothesis. *Applied Linguistics*, 24(2), 168-196.
- Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the Output Hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition*, 21, 421-452.
- James, C. (1974). The consistency of marking a physics examination. *Physics Education*, 9, 271-274.
- James, C. (1977). Judgments of error gravities. *ELT Journal*, 31(2), 116-124.
- James, W. (1890). *The Principles of psychology*. New York: Holt.
- Jamieson, J. (2004). Analysis of covariance (ANCOVA) with difference scores. *International Journal of Psychophysiology*, 52, 277-283.
- Jin, L., & Cortazzi, M. (2006). Changing practices in Chinese cultures of learning. *Language, Culture and Curriculum*, 19(1), 5-20.
- Johnson, M. D. (2011). *Planning in second language writing*. PhD Thesis. Northern Arizona University.
- Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric *Applied Measurement in Education*, 13(2), 121-138.
- Johnson, R. L., Penny, J., & Gordon, B. (2001). Score resolution and interrater reliability of holistic scores in essay ratings. *Written Communication*, 18, 229-249.
- Jones, S., & Tetroe, J. (1987). Composing in a second language. In A. Matsushashi (Ed.), *Writing in real time: Modelling production processes* (pp. 34-57). Norwood, NJ: Ablex Publishing Corporation.
- Karmiloff-Smith, A. (1986). Stage/structure versus phase/process in modelling linguistic and cognitive development. In I. Levin (Ed.), *Stage and structure: Reopening the debate*. Norwood: N.J.: Ablex.
- Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 142-166). Amsterdam: John Benjamins Publishing Company.
- Kellogg, R. T. (1987). Writing performance: Effects of cognitive strategies. *Written Communication*, 4(3), 269-298.
- Kellogg, R. T. (1988). Attentional overload and writing performance: effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(2), 355-365.
- Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*, 103(3), 327-342.
- Kellogg, R. T. (1996). A model of working memory in writing. In C. Levy, & S. Ransdell (Eds.), *The science of writing* (pp. 57-71). Mahwah, NJ.: Erlbaum.

- Kinncar, P. R., & Gray, C. D. (2009). *SPSS 16 made simple*. East Sussex: Psychology Press.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review* 85, 363-394.
- Knoch, U. (2007). *Diagnostic writing assessment: The development and validation of a rating scale*. PhD thesis. The University of Auckland.
- Kormos, J. (2006). *Speech production and second language acquisition*. NJ: Mahwah: Lawrence Erlbaum Associates.
- Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20, 148-161.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, 73, 440-464.
- Krippendorff, K. (2005). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks ; London ; New Delhi: Sage.
- Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing* 17, 48-60.
- Lantolf, J. P., & Pavlenko, A. (2001). (S)econd (L)anguage (A)ctivity theory: understanding second language learners as people. In M. P. Breen (Ed.), *Learner contributions to language learning: New directions in research* (pp. 141-158). Harlow, Essex: Pearson Education.
- LaPierre, D. (1994). *Language output in a cooperative learning setting: Determining its effects on second language learning* (PhD Thesis). University of Toronto.
- Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Lee, I. (2002). Helping students develop coherence in writing. *English Teaching Forum*, July, 32-39.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: The MIT Press.
- Li, D. (2004). *The effects of pre-task planning and online planning on learners' written language performance*. MA Thesis. The University of Auckland. Auckland.
- Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, 32, 169-192.
- Li, P., Jin, Z., & Tan, L. (2004). Neural representations of nouns and verbs in Chinese: An fMRI study. *NeuroImage*, 21(4), 1533-1541.
- Li, Z. (2008). Quantitative analysis of the effects of English majors' planning variables on their writing performance. *Foreign Language Teaching and Research*, 40(3), 178-183.
- Liang, M. (2006). A study of coherence in EFL learners' written production. *Modern*

*Foreign Language*, 29(3), 284-292.

- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(126-141).
- Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language training. In K. H. M. Pienemann (Ed.), *Modelling and assessing second language acquisition* (pp. 77-99). Clevedon: Multilingual Matters.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie, & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.
- Long, M. H. (2000). Focus on form in task-based language teaching. In R. D. Lambert, E. G. Shohamy, & A. R. Walton (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton*. Amsterdam: John Benjamins Publishing Co.
- Long, M. H., & Crookes, G. (1991). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26, 27-55.
- Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. . In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (Vol. 15-41). Cambridge: Cambridge University Press.
- Louwerse, M. M. (2002). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, 12, 291-315.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36-62.
- Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221.
- Lynch, T., & Maclean, J. (2001). 'A case of exercising': Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 141-162). Harlow: Pearson Education Limited.
- Ma, G. (2002). Contrastive analysis of linguistic features between EFL and ENL essays. *Foreign Language Teaching and Research*, 34(5), 345-349.
- Malvern, D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Manchon, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing. *Language Learning*, 57, 549-593.
- Mathews, R. C., Buss, R. R., Stanley, W. B., Blachard-Fields, E., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1083-1100.
- McCarthy, M. (2007). *Assessing development of advanced proficiency through learner corpora*. manuscript.
- McLaughlin, B. (1987). *Theories of second language learning*. London: Edward Arnold.
- McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11, 113-128.
- McLaughlin, B., & Heredia, R. (1996). Information-processing approaches to research on

- second language acquisition. In W. Ritchie, & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 213-228). San Diego: Academic Press.
- McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information processing perspective. *Language Learning*, 33, 135-158.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1-43.
- McNamara, D. S., Louwrese, M. M., Cai, Z., & Graesser, A. (2005). Coh-Metrix version (Version 1.4). Retrieved from <http://cohmetrix.memphis.edu>.
- McNamara, D. S., Louwrese, M. M., & Graesser, A. C. (2002). *Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension*. Institute for Intelligent Systems, University of Memphis, Memphis, TN.
- McVey, P. J. (1975). The errors in marking examination scripts in electronic engineering. *International Journal of Electronic Engineering Education*, 12, 203-216.
- Meara, P. (2001). P-Lex (Version 2.0) [Computer program].
- Meara, P., & Bell, H. (2001). P-Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5-19.
- Meara, P., & Miralpeix, I. (2004a). D-Tool [Computer program].
- Meara, P., & Miralpeix, I. (2004b). *D-Tool-Manual*. manuscript.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20(01), 83-108.
- Mehrang, F., & Rahimpoura, M. (2010). The impact of task structure and planning conditions on oral performance of EFL learners. *Procedia Social and Behavioral Sciences*, 2, 3678-3686.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, 12(1), 11-37.
- Nakakubo, T. (2011). *The effects of planning on second language oral performance in Japanese: Processes and production* (PhD Thesis). The University of Iowa.
- Nation, I. S. P. (1988). *Word lists: Words, affixes and stems*. Wellington: Victoria University of Wellington Press.
- . *National English Curriculum Standards*. (2001).
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: NJ: Prentice-Hall, INC.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578.



- O'Malley, J., & Chamot, A. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Ojima, M. (2006). Concept mapping as pre-task planning: A case study of three Japanese ESL writers. *System*, 34(4), 566-585.
- Olson, D. R. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Education Review*, 47, 257-281.
- Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, 19, 218-233.
- Ortega, L. (1995a). *The effect of planning on L2 Spanish narratives*. Honolulu HI: University of Hawaii, Second Language Teaching and Curriculum Center.
- Ortega, L. (1995b). *Planning and second language oral performance*. MA thesis. University of Hawai'i. Honolulu.
- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21(01), 109-148.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518.
- Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77-109). Amsterdam: John Benjamins Publishing Company.
- Ortega, L. (2007). *Social context in task-based language learning: (How) does it matter?* Paper presented at the meeting of Conference on Social and Cognitive Aspects of Second Language Learning and Teaching.
- Philp, J., Oliver, R., & Mackey, A. (2006). The impact of planning time on children's task-based interactions. *System*, 34, 547-565.
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, 10, 52-79.
- Pienemann, M. (1998). *Language, processing and second language development: Processability theory*. Amsterdam: John Benjamins.
- Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.
- Pu, X. (2009). Effects of prewriting discussion on the language quality of argumentative writings. *Teaching English in China*, 32(5), 16-24.
- Quirk, R., & Greenbaum, S. (1973). *A university grammar of English*. London: Longman Press.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A grammar of contemporary English*. London: Longman Group Ltd.
- Rahimpour, M., & Safarie, M. (2011). The effects of on-line and pre-task planning on descriptive writing of Iranian EFL learners. *International Journal of English Linguistics*, 1(2), 274-280.
- Rau, P., & Sebrechts, M. (1996). How initial plans mediate the expansion and resolution of options in writing. *The Quarterly Journal of Experimental Psychology*, 49, 616-638.

- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 119-135.
- Reid, J. (1992). A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. *Journal of Second Language Writing*, 1(2), 79-107.
- Rijlaarsdam, G., & van den Bergh, H. (2006). Writing process theory: A functional dynamic approach. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 41-53). New York: Guilford.
- Risemberg, R. (1996). Reading to write: Self-regulated learning strategies when writing essays from sources. *Literacy Research and Instruction*, 35(4), 365-383.
- Roberts, R., & Kreuz, R. J. (1993). Nonstandard discourse and its coherence. *Discourse Processes*, 16, 451-464.
- Robinson, P. (1995). Attention, memory, and the 'Noticing Hypothesis'. *Language Learning*, 45, 283-331.
- Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, 18, 27-67.
- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.
- Robinson, P. (2003). The Cognition Hypothesis, task design, and adult task-based language learning *Second Language Studies*, 21(2), 45-105.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching (IRAL)*. 43(1), 1-32.
- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, intention, uptake, and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching (IRAL)*. 45(3), 193-213.
- Robinson, P. (2011). *Second language task complexity*. Amsterdam: John Benjamins Publishing Company.
- Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *IRAL*, 45, 161-176.
- Roebuck, R. (2000). Subjects speak out: How learners position themselves in a psycholinguistic task. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 79-96). Oxford, NY: Oxford University Press.
- Roediger, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, 2, 123-127.
- Romanova, N. (2010). Planning, recast, and learning of L2 morphology. *The Canadian Modern Language Review*, 66(6), 843-875.
- Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

- Rutherford, K. (2001). *An investigation of the effects of planning on oral production in a second language*. MA Thesis. The University of Auckland. Auckland.
- Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Basingstoke: Palgrave Macmillan.
- Sangarun, J. (2001). *The effects of pre-task planning on foreign language performance* (PhD dissertation). University of Toronto, Toronto.
- Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111-141). Amsterdam: John Benjamins Publishing Company.
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing* 9(3), 259-291.
- Schmidt, R. (1990). The role of consciousness in second language learning *Applied Linguistics*, 11, 129-158.
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357-385.
- Schmidt, R. (1995). *Attention and awareness in foreign language learning* (9). Hawaii: University of Hawaii Second Language Teaching and Curriculum Center.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). New York: Cambridge University Press.
- Shi, L. (1998). Effects of prewriting discussions on adult ESL students' compositions. *Journal of Second Language Writing*, 7(3), 319-345.
- Shiffrin, R. M., & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 111-140). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84(2), 127-190.
- Shin, Y. (2008). *The effects of planning on L2 writing: A study of Korean learners of English as a foreign language*. PhD Thesis. The University of Iowa.
- Simon, H. A. (1978). Information-processing theory of human problem solving. In W. K. Estes (Ed.), *Handbook of learning and cognition processes* (pp. 271-295). Hillsdale, N.J: Erlbaum Associates.
- Skehan, P. (1996a). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38-62.
- Skehan, P. (1996b). Second language acquisition research and task-based instruction. In J. Willis, & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17-30). Bath: Heinemann.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2000). Re-interpreting focus-on-form: The role of tasks and teachers. *EA Journal*, 18(2), 8-20.
- Skehan, P. (2009a). Lexical performance by native and non-native speakers on language learning tasks. In B. J. Richards, H. Daller, D. Malvern, P. Meara, J. Milton, & J.

- Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 107-126). London: Palgrave Macmillan.
- Skehan, P. (2009b). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532.
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1(3), 185.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93-120.
- Skehan, P., & Foster, P. (2005). Strategic and on-line planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193-216). Amsterdam: John Benjamins Publishing Company.
- Skibniewski, L., & Skibniewska, M. (1986). Experimental study: The writing processes of intermediate/advanced foreign language learners in their foreign and native languages. *Studia Anglica Posnaniensia*, 19, 142-163.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, 50, 158-164.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). Cambridge: Cambridge University Press.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371-392.
- Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, 82(3), 320-337.
- Tajima, M. (2003). *The effects of planning on oral performance of Japanese as a foreign language*. PhD dissertation. Purdue University.
- Tannen, D. (1985). Relative focus on involvement in oral and written discourse. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning* (pp. 124-147). Cambridge: Cambridge University Press.
- Tarone, E. (1979). Interlanguage as chameleon. *Language Learning*, 29, 181-192.
- Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, 4(2), 142-164.

- Tarone, E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, 35, 373-403.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-273). Amsterdam: John Benjamins Publishing Company.
- Taylor, K., & Dionne, J. (2000). Assessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, 92(3), 413-425.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84-119.
- van den Bergh, H., & Rijlaarsdam, G. (1999). The dynamics of generation during writing: An online study. In M. Tonance, & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 99-118). Amsterdam: Amsterdam University Press.
- van der Hoeven, J. (1999). Differences in writing performance: Generating as an indicator. In M. Tonance, & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 63-77). Amsterdam: Amsterdam University Press.
- Vann, R. J., Meyer, D. E., & Lorenz, E. O. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly*, 18(3), 427-440.
- VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12.
- VanPatten, B. (1996). *Input processing and grammar instruction*. New York: Ablex.
- Vygotsky, L. S. (1962). *Thought and language* (E. Haufmann, & G. Vakar, Trans.). Cambridge, MA: MIT Press.
- Wang, Z. J. (2009). *Modeling L2 speech production and performance: Evidence from five types of planning and two task structures*. (PhD thesis). The Chinese University of Hong Kong, Hong Kong.
- Watson Todd, R., Thienpermpool, P., & Keyuravong, S. (2004). Measuring the coherence of writing using topic-based analysis. *Assessing writing*, 9, 85-104.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wendel, J. N. (1997). *Planning and second language narrative production*. PhD dissertation. Temple University.
- Whalen, K., & Menard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning*, 45(3), 381-418.
- Wickens, C. D. (1989). Attention and skilled performance. In D. H. Holding (Ed.), *Human skills* (pp. 71-105). Chichester: John Wiley & Sons.
- Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14(1), 85-106.
- Wigglesworth, G., & Elder, C. (2010). An investigation of the effectiveness and validity of planning time in speaking test tasks. *Language Assessment Quarterly*, 7(1), 1-24.

- Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly*, 26(4), 693-709.
- Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford: Oxford University Press.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity* (17). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Yuan, F. (2001). *The effects of planning on language production in task-based language teaching*. Doctoral Dissertation. Temple University.
- Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24(1), 1-27.
- Yule, G. (1996). *The study of language* (2nd ed.). Cambridge: Cambridge University Press.
- Zimmerman, R. (2000). L2 writing: Subprocesses, a model of formulating and empirical findings. *Learning and Instruction*, 10, 73-99.

APPENDIX A: CONSENT FORM

**The Consent Form (Student)**

*This consent form will be held for a period of six years.*

**Project title: The impact of writing practice on language learning**

Principal investigator: Yan Lin

I have been given and have understood an explanation of this research project. I have had an opportunity to ask questions and all my questions have been answered to my satisfaction.

- I understand that my participation in this research is voluntary. I also understand that if I choose not to participate, I can withdraw from the study any time during the research period. I understand that my teacher has given assurance that my participation or nonparticipation or withdrawal will not affect my relationship with the teacher or my grades.
- I understand that I will be asked to give 5 hours outside of class time to this research.
- I understand that my answers to interview questions will be audio- recorded, transcribed and coded. The audio recording can be stopped on my request at any time during the process.
- I understand that I will not be given an opportunity to edit the transcripts of the audio- recordings of my answers. Nor will I be offered a copy of the electronic files on my answers and written production.
- I understand that my written production, written notes and answers to questions in the questionnaires will be collected and coded for analysis by the researcher. I will not be offered an opportunity to edit my written production, written notes or answers to questionnaire questions.
- I understand that my test results will be collected and analyzed by the researcher.
- I understand that my name and the name of my university will not be used in any form of reporting of this research.
- I understand that I have the right to withdraw any information traceable to me up to 31/07/2010 without giving a reason.
- I understand that a small gift will be given as a token of gratitude for my participation.
- I agree to participate in the study.
- I would /would not like to have a copy of the study report. (Please circle)
- I agree/disagree to be audio recorded. (Please circle)
- I agree / disagree to the use of my written production and audio clips of my answers to the questions in research presentations. (Please circle)

\_\_\_\_\_  
Name of the participant

\_\_\_\_\_  
Signature of the participant

\_\_\_\_\_  
Date

**APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON .....for ...(3).....years on ...(date)..., Reference Number 2009/.**

## APPENDIX B: QUESTIONNAIRE FOR BACKGROUND INFORMATION

Please complete the following information sheet. All the data gathered will be used only as background information for the present study. All the information provided on this form will be kept strictly confidential. Thank you for your cooperation.

1. Name \_\_\_\_\_ Department \_\_\_\_\_ Class \_\_\_\_\_

2. Gender:    Male     Female

3. Age \_\_\_\_\_

4. Did you study English at \_\_\_\_\_? For how long ?

(Please put a tick by the type of school you choose, then indicate how many years you studies English there. For example: Elementary school  for 3 years.)

Elementary school  for \_\_\_\_\_ years

Middle School  for \_\_\_\_\_ years

High School  for \_\_\_\_\_ years

5. Have you ever traveled \_\_ or lived \_\_ in an English speaking country?

Yes     No

If 'yes',

a. Which country? (please specify) \_\_\_\_\_

b. For how long? \_\_\_\_\_

6. Do you speak English after class? If yes, please specify how frequent and with whom (e.g. native speakers of English, your classmates) you speak English.

7. Do you listen to English broadcast such as VOA and BBC? If yes, how often?

8. Do you read in English beyond the English textbooks? If yes, please specify what you often read.

9. Do you attend night or weekend English class? If yes, please specify how often you attend such a class and what class activities you usually have.

Test score (for the researcher to fill out)

\_\_\_\_\_



## APPENDIX C: WRITING TASKS FOR THE TESTING SESSIONS

1. You are a high school graduate and have got offers from a good university in China and a university in UK. You want to go to that university in UK but your parents want you to go to the university in China. You can make use the following information and other reasons you can think of to persuade your parents by **comparing the two universities**.

**You will have limited time to write. Please try to use different adjectives when you are comparing the two universities. Please write as much as possible. It is best if you could write a page.**

	The Chinese university	The university in UK
World ranking	225	100
Tuition	¥ 8000=about £ 800	£ 9,975 pounds
Teaching staff	80% of staff has PhDs from some of the best universities in the world.	Staff has PhDs from some of the best universities in the world and are among the top ranking researchers
Accommodation	4-year on campus dormitories	First year on campus The rest of the time off campus
Living expenses	¥ 1000=about £ 100 /per month	£ 784 /per month
Study options	Full time	Full time as well as part-time studies
Financial support	¥ 20,000=around £ 2000 /per year scholarship	£ 2,500 pounds /per year scholarship
Years of study	4	3
Facilities	Library holding 1.8 million items. Gym with in-door swimming pool	Library holding more than 2.2 million items, many of which can be viewed via the impressive online library system.
Opportunity to prepare for your future career		
Opportunity to travel		
Opportunity to learn English and British culture		
Opportunity to be independent		
Other		

**Here are some words and phrases you might need in your composition:** cheap, expensive, qualified, free, convenient, flexible, supportive, impressive, options, eye-opening, open-minded, knowledgeable, understanding.

2. Your cousin, who is also a good friend of yours, has written to you and told you that s/he is not enjoying high school and so is thinking of not going to University. You are at University and you are really looking forward to your cousin being there with you. Write a letter to your cousin. In your letter you will try to have your cousin see that University is different from High school in many aspects and your life at University is fantastic. You need to persuade him/her to join you next year.

The following table may help you compare University with High school. You will have limited time to write. Please try to use different adjectives when you are comparing University with High School. Please write as much as possible. It is best if you could write a page.

	University	High school
Classroom hours	18 hours/week	40 hours/week
Teachers' qualification	All have MA or higher degrees	All have BA degrees and some have MA degrees
Living conditions	Live in dormitories on campus; Eat in university canteens	Live at home; Eat at home
Free time	More than 4 hours/per day	About 3 hours/per day
Extracurricular activities	All kind of clubs Cultural festivals All kinds of contests, etc.	
Freedom		
Choice of subjects		
Social opportunities		
Attitude of teachers to students		
Opportunities to develop practical skills		

**Here are some words and phrases you might need for your composition: heavy workload, highly qualified, convenient, colorful life, free, practical, sociable, friendly, approachable, prepare for the future**

3. You are English major who is graduating from the university and is looking for a job. You have two job offers. Below are the job descriptions. Your parents want you to take Job 1 but you would prefer Job 2. Write a letter to your parents explaining why you would prefer Job 2.

You will have limited time to write. Please try to use different adjectives when you are comparing the two jobs. Please write as much as possible. It is best if you could write a page.

	Job 1 English Instructor in a high school	Job 2 Marketing in a big trading company
Salary	¥3000/per month	¥ 5000/per month
Working hours	20 hours/week	40 hours/week
Vacation	2 months/year	15 days/year
Opportunities to travel	few	many
Job stability	high	Medium--low
Training opportunities	A few	many
Responsibilities	<ol style="list-style-type: none"> <li>1. Teach English to three classes of students.</li> <li>2. Have 6 hours of office hour to help students.</li> <li>3. Attend staff meetings.</li> </ol>	<ol style="list-style-type: none"> <li>1. Collect and analyze information on the market and target customers</li> <li>2. Maintain customer relationship.</li> <li>3. Market commodities.</li> </ol>

**Here are some words and phrases you might need for your composition:** stable, challenging, tiring, interesting, difficult, sociable, responsible, flexible, relax, energetic, develop career.

## APPENDIX D: WRITING TASKS FOR THE TREATMENT SESSIONS

1. Your friend has just given birth to a baby girl. She is quite unhappy since she wanted very much to have a baby boy. As her friend, you are trying to cheer her up by telling her why you think a girl is better than a boy. Write a letter to her to **compare a girl with a boy**. You may consider the following aspects when comparing. Try to think of the differences between a boy and a girl at different ages. Please give examples to illustrate your points.



**You will have limited time to write. Please try to use different adjectives when you are comparing a boy with a girl. Please write as much as possible. It is best if you could write a page.**

- The attention usually needed by a boy/girl
- The money spent on a boy/girl
- The help parents get from a boy/girl
- The attention and company parents get from a boy/girl

**Here are some adjectives you might need in your composition:** naughty, clean, expensive, careful, caring, helpful, sensitive, pretty, strong, brave, clever, hard to control.

2. A travel agency is looking for a new travel agent. Now there are two candidates for the job. As one of the recruiters, decide which one you would like to choose and try to convince other people in the selection panel that the person you choose is the right one. You can use the information below to **compare the person you choose with the other candidate**. You can also use other information from your world knowledge to justify your choice.

**You will have limited time to write. Please try to use different adjectives when you are comparing the two candidates. Please write as much as possible. It is best if you could write a page.**

	 Dave	 Jacqueline
Age	35	45
Work experience	5 years salesman 5 years IT assistant	3 years supermarket cashier 5 years receptionist 5 years school teacher
Education background	BA in Business	BA in Math MA in Psychology
Personality	Out-going, quick-tempered	Organized, caring

Asking salary	\$ 50,000	\$ 45,000
Marital status	Married but no children	Married with three children
Other information	Has traveled a lot; Has excellent references;	Pretty; Has not traveled much; Has good social skills

**Here are some words and phrases you might need in your composition:** experienced, qualified, suitable, related, sociable, heavy burden, capable, likely to gain trust.

3. Your parents are going to buy you an apartment. Below is some information on two apartments. They want to buy Apartment 2 but you want them to buy Apartment 1. Write them a letter explaining why your choice is superior to theirs.

**You will have limited time to write. Please try to use different adjectives when you are comparing the two apartments. Please write as much as possible. It is best if you could write a page.**

	Apartment 1	Apartment 2
Location	10km from city center	40 km from city center, in a suburb
Size	120 square meters	140 square meters
Transportation	10 minutes' walk to buses, subway stations	8 minutes' walk to Bus 890 that can take you to the subway station
Price	¥1,2000/per square meter	¥8000/per square meter
Level	Level 2 in a 12-story building	Level 15 in a 20-story building
Rooms	2 bedrooms, 1 living room, 1 kitchen, 2 bathrooms	3 bedrooms, 2 living rooms, 1 kitchen, 2 bathrooms
Amenity and facilities	5 min walk to shopping mall with a cinema and a gym, 10 min drive to a hospital, 10 min walk to schools	10 min walk to a supermarket, 20 min drive to a hospital, 20 min drive to a cinema, 30 min walk to schools
Neighborhood	All kinds of people	Mostly white-collar workers, professionals, university professors

**Here are some words and phrases you might need for your composition:** close, convenient, cheap, expensive, bright, easy to get to..., safe, suitable, comfortable, far, get along.

4. You have a pen pal from country X. From his/her letter you find that s/he does not have much knowledge of what your country is like and how your life is now. S/he thinks you are living like how you were 10 years ago. Write a letter to tell him/her how your life is different from 10 years ago. Try to write as many aspects and details as you can. The following aspects are just for your reference. You can think of other aspects to include when you are comparing your present life with that 10 years ago.

**You have limited time to write. Please try to use different adjectives when you are comparing. Please write as much as possible. It is best if you could write a page.**

- People's income
- Living conditions
- Variety of commodities in the shops
- Price of commodities, e.g. food, clothing, houses, electronic device etc.
- Education
- Access to information
- Environment, e.g. air quality, water, etc.
- People's life span

**Here are some words and phrases you might need for your composition:** salary, improve, various, wide range of..., polluted, heavy pollution, healthy.

## APPENDIX E: POST-TASK QUESTIONNAIRE

### *Questionnaire for the pre-task planning group*

Please answer the following questions as truthfully as possible.

1. Name\_\_\_\_ Grade\_\_\_\_ Department\_\_\_\_\_
2. Please explain in detail how you planned. (How did you spend your time? How much time did you actually use in preparing for the writing? What did you do during the preparation time?)
3. What would you say your focus was when you were preparing to write?
4. Did you encounter any difficulties? If 'yes', what were they? How did you overcome the difficulties?
5. When you were writing did you try to\_\_\_\_\_? (Please check the ones that are true to you. You can choose more than one option.)
  - a. keep it simple
  - b. write much
  - c. use language you are sure of
  - d. try to express complex ideas with language you are not sure of

For question 6, 7 & 8, please give your answer by circling one of the numbers.

6. Did you use much of the content you had planned in your writing?

None All  
1          2          3          4          5

Did you use much of the language you had planned in your writing?

None All  
1          2          3          4          5

7. Did you find the tasks difficult?

Easy Difficult  
1          2          3          4          5

8. Did you feel pressed for time when doing the task?

No Yes  
1          2          3          4          5

9. Did you have time to think of the vocabulary and grammar needed for expression?

Vocabulary  
No Yes  
1          2          3          4          5

Grammar  
No Yes  
1          2          3          4          5

10. Did you pay attention to accuracy when you were writing?

No				Yes
1	2	3	4	5

11. How much time approximately did you spend on organizing ideas during writing?

12. Did you have time to edit your essay before handing it in? If 'yes', how long?

13. How did you feel about having time to plan? Did it help?

No				Yes
1	2	3	4	5

In what ways?

14. If you could choose, would you choose to have time to prepare before starting to write or start writing right after you read the instructions?

Have time to prepare       Start writing right after reading the instructions



*Questionnaire for the no planning group*

Please answer the following questions as truthfully as possible.

1. Name\_\_\_\_ Grade\_\_\_\_ Department\_\_\_\_\_
2. Did you encounter any difficulties? What were they? How did you overcome them?
3. When you were writing did you try to\_\_\_\_\_? (Please check the ones that are true to you. You can choose more than one option.)
  - a. keep it simple
  - b. write much
  - c. use language you are sure of
  - d. try to express complex ideas with language you are not sure of

For question 4, 5 & 6, please circle a number that reflects your true feeling.

4. Did you find the tasks difficult?

Easy				Difficult
1	2	3	4	5

5. Did you feel pressed for time when doing the task?

No				Yes
1	2	3	4	5

6. Did you have time to think of the vocabulary and grammar needed for expression?

Vocabulary				
No				Yes
1	2	3	4	5
Grammar				
No				Yes

7. Did you pay attention to accuracy when you were writing?

No				Yes
1	2	3	4	5

8. How much time approximately did you spend on organizing ideas?

9. Did you have time to edit your essay before handing it in? If 'yes', how long?

10. How did you feel about having 15 minutes to finish writing without any time to prepare in advance?

11. If you can choose, would you choose to have time to prepare before starting to write or start writing right after you read the instructions?

Have time to prepare       Start writing right after reading the instructions

*Questionnaire for the online planning group*

Please answer the following questions as truthfully as possible.

1. Name\_\_\_\_ Grade\_\_\_\_ Department\_\_\_\_\_

2. Did you encounter any difficulties? What were they? How did you overcome them?

3. When you were writing did you try to\_\_\_\_\_? (Please check the ones that are true to you. You can choose more than one option.)

- a. keep it simple
- b. write much
- c. use language you are sure of
- d. try to express complex ideas with language you are not sure of

For question 4, 5 & 6, please circle a number that reflects your true feeling.

4. Did you find the tasks difficult?

Easy					Difficult
1	2	3	4	5	

5. Did you feel pressed for time when doing the task?

No				Yes
1	2	3	4	5

6. When you were writing did you have time to think of vocabulary and grammar needed for expression?

Vocabulary				
No				Yes
1	2	3	4	5

Grammar				
No				Yes
1	2	3	4	5

7. Did you pay attention to accuracy when you were writing?

No				Yes
1	2	3	4	5

8. How much time approximately did you spend on organizing ideas?

9. Did you have time to edit your essay before handing it in? If 'yes', how long?

10. How did you feel about having 25 minutes to finish writing without any time to prepare in advance?

11. If you can choose, would you choose to have time to prepare before starting to write or start writing right after you read the instructions?

Have time to prepare

Start writing right after reading the instructions

## APPENDIX F: POST-TASK INTERVIEW QUESTIONS

### *Questions for the pre-task planning group*

1. Have you ever been taught how to prepare for writing?
2. Did you prepare in the same way or differently for different topics?
3. Could you please describe your planning procedure?
4. What would you say your focus was when you were planning?
5. Did you spend on thinking about content? Follow-up: How much time did you spend on it?
6. Did you think of grammar and vocabulary? Possible follow-up question: How much time did you spend on thinking about vocabulary? How about sentence patterns? Grammar?
7. When you were planning, were you thinking in English or in Chinese?
8. Were you able to apply what you prepared in your writing? Follow-up: How much of your plan was used in your writing? Which part of your plan did you use much in your writing, content or language? Why couldn't you use much of the planned (content or language) in your writing?
9. What do you think you were focusing on when you wrote?
10. When you were writing did you have much time to think of grammar and vocabulary?
11. Did you pause during writing?
12. When you paused during writing what were you thinking?
13. Did you ever go back and read what you had written? Why did you do that?
14. Did you have time to edit your essay?
15. Did you manage to correct some mistakes? What kind of mistakes?
16. Please comment on the fact that you had 10 minutes to plan before writing?
17. Did you use up the 10 minutes?
18. If you could choose, would you choose to have time to prepare or not? Why/Why not?
19. If you were given the same type of task to do again, would you plan in the same way or differently? Why?

### *Questions for the online planning group*

1. Please tell me how you made use of the 25-minute writing time.
2. What do you think you were focusing on when you wrote?
3. When you were writing did you have much time to think of grammar and vocabulary? Follow-up: How much time approximately did you spend on thinking of vocabulary? Grammar?
4. Did you pause during writing?
5. When you paused during writing what were you thinking?
6. Did you ever go back and read what you had written? Why did you do that?
7. Did you have time to edit your essay?
8. Did you manage to correct some mistakes? What kind of mistakes?
9. If you were given a chance to write this again, would you write differently?
10. How did you feel about having to write without any time to prepare?
11. If you could choose, would you choose to have time to prepare before starting to write or start writing right after you read the instructions?

### *Questions for the no planning group*

- 1 Please tell me how you made use of the 15-minute writing time.
2. What do you think you were focusing on when you wrote?
3. When you were writing did you have much time to think of grammar and vocabulary? Follow-up:  
How much time approximately did you spend on thinking of vocabulary? Grammar?
4. Did you pause during writing?
5. When you paused during writing what were you thinking?
6. Did you ever go back and read what you had written? Why did you do that?
7. Did you have time to edit your essay?
8. Did you manage to correct some mistakes? What kind of mistakes?
9. If you were given a chance to write this again, would you write differently?
10. How did you feel about having to write without any time to prepare?
- 11.If you could choose, would you choose to have time to prepare before starting to write or start writing right after you read the instructions?

## APPENDIX G: POST-STUDY QUESTIONNAIRE

Please answer the following questions truthfully. Thank you !!!

1. How do you feel about writing in English in general? (e.g. Do you like it or not? Do you feel nervous?)
2. How do you feel about making mistakes? Did you try to be accurate when you wrote?
3. Did you have enough time to finish the task?  
Was the time too long\_\_\_\_, too short\_\_\_\_, or just right\_\_\_\_\_?
4. Did you do similar tasks during the past five weeks (e.g. read, speak or write on a similar topic)? If 'yes', please specify.
5. Did you have instructions on comparative forms in this period? Yes  No
6. Did you have instructions on other grammar rules? If 'yes', what were they?
7. What do you think the focus of the study was?

## APPENDIX H: COH-METRIXINDEXES

No.	Description	Full description
	Title	Title
1	Causal content	Incidence of causal verbs, links, and particles
2	Causal cohesion	Ratio of causal particles to causal verbs (cp divided by cv+1)
3	Pos. additive connectives	Incidence of positive additive connectives
4	Pos. temporal connectives	Incidence of positive temporal connectives
5	Pos. causal connectives	Incidence of positive causal connectives
6	Neg. additive connectives	Incidence of negative additive connectives
7	Neg. temporal connectives	Incidence of negative temporal connectives
8	Neg. causal connectives	Incidence of negative causal connectives
9	All connectives	Incidence of all connectives
10	Adjacent argument overlap	Argument Overlap, adjacent, unweighted
11	Adjacent stem overlap	Stem Overlap, adjacent, unweighted
12	Adjacent anaphor reference	Anaphor reference, adjacent, unweighted
13	Argument overlap	Argument Overlap, all distances, unweighted
14	Stem overlap	Stem Overlap, all distances, unweighted
15	Anaphor reference	Anaphor reference, all distances, unweighted
16	NP incidence	Noun Phrase Incidence Score (per thousand words)
17	Pronoun ratio	Ratio of pronouns to noun phrases
18	Conditional operators	Number of conditional expressions, incidence score
19	Negations	Number of negations, incidence score
20	Logic operators	Logical operator incidence score (and + if + or + cond + neg)
21	LSA sentence adjacent	LSA, Sentence to Sentence, adjacent, mean
22	LSA sentence all	LSA, sentences, all combinations, mean
23	LSA paragraph	LSA, Paragraph to Paragraph, mean
24	Personal pronouns	Personal pronoun incidence score
25	Noun hypernym	Mean hypernym values of nouns
26	Verb hypernym	Mean hypernym values of verbs
27	No. of paragraphs	Number of Paragraphs
28	No. of sentences	Number of Sentences

29	No. of words	Number of Words
30	Sentences per paragraph	Average Sentences per Paragraph
31	Words per sentence	Average Words per Sentence
32	Syllables per word	Average Syllables per Word
33	Flesch Reading Ease	Flesch Reading Ease Score (0-100)
34	Flesch-Kincaid	Flesch-Kincaid Grade Level (0-12)
35	Modifiers per NP	Mean number of modifiers per noun-phrase
36	Higher level constituents	Mean number of higher level constituents per word
37	Words before main verb	Mean number of words before the main verb of main clause in sentences
38	Type-token ratio	Type-token ratio for all content words
39	Raw freq. content words	Celex, raw, mean for content words (0-1,000,000)
40	Log freq. content words	Celex, logarithm, mean for content words (0-6)
41	Min. raw freq. content words	Celex, raw, minimum in sentence for content words (0-1,000,000)
42	Log min. freq. content words	Celex, logarithm, minimum in sentence for content words (0-6)
43	Concreteness content words	Concreteness, mean for content words
44	Pos. logical connectives	Incidence of positive logical connectives
45	Neg. logical connectives	Incidence of negative logical connectives
46	Intentional cohesion	Ratio of intentional particles to intentional content
47	Intentional content	Incidence of intentional actions, events, and particles.
48	Temporal cohesion	Mean of tense and aspect repetition scores
49	Syntactic structure similarity adjacent	Sentence syntax similarity, adjacent
50	Syntactic structure similarity all-1	Sentence syntax similarity, all, across paragraphs
51	Syntactic structure similarity all 2	Sentence syntax similarity, sentence all, within paragraphs
52	Content word overlap	Proportion of content words that overlap between adjacent sentences
53	Spatial cohesion	Mean of location and motion ratio scores.
54	Min. concreteness content words	Concreteness, minimum in sentence for content words
55	GNRPure	Genre purity



## APPENDIXI: SLA STUDIES ON PRE-TASK PLANNING USING ORAL PRODUCTION DATA

Study	Learners and context	Time span of the main study	Research questions or hypotheses	Method and instruments	Measures	Results (quantitative)
Crookes (1989)	40 Japanese learners of English in the University of Hawaii/ESL	Less than 1 day (two oral tasks with a questionnaire for biodata information in between)	Does planning lead to speech which differs from that produced without planning in ways relevant to second language development?	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Two monologic oral tasks: a lego task and a map task.</li> <li>➤ Two planning conditions : no planning; pre-task planning</li> <li>➤ Planning time: 10 minutes</li> </ul>	<p>Syntax:</p> <ul style="list-style-type: none"> <li>● Words/utterance</li> <li>● Subordination/T-unit</li> <li>● Subordination/utterance</li> <li>● S-nodes/utterance</li> <li>● words/subordination</li> <li>● VP range/utterance</li> </ul> <p>Lexis:</p> <ul style="list-style-type: none"> <li>● Type-token ratio (TTR);</li> <li>● Adjective sequence/utterance;</li> <li>● Words/adjective sequence;</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Words/error-free T-unit;</li> <li>● Error-free T-unit/utterance;</li> </ul>	<ol style="list-style-type: none"> <li>1. Planning produced more complex language as measured in terms of words per utterance, number of subordinate clauses per utterance, and s-nodes per utterance.</li> <li>2. Planning effects were also found in lexical variety (TTR) but not in syntactic variety (VP range per utterance).</li> <li>3. No statistical significant increase in accuracy was found in all the measures except TLU <i>the</i> in task 1.</li> <li>4. In one of the tasks NNSs produced more discourse markers under the planning condition.</li> </ol>

					<ul style="list-style-type: none"> <li>● Target-like usage (TLU) of plural -s&amp; concord;</li> <li>● TLU <i>the</i>;</li> <li>● TLU <i>a</i></li> </ul> Discourse: <ul style="list-style-type: none"> <li>● Discourse markers/utterance</li> </ul>	
Williams (1992)	24 NNS teaching assistants (TA) in an American university/ESL	2 years(Two presentation sessions of about 7-8 minutes each for each participant. The two sessions were two weeks apart.)	1. Would planned production contain more overt marking of discourse functions than unplanned production? 2. Would comprehensibility increase with more overt marking of discourse Functions? 3. Does syntactic or morphological accuracy and complexity cause	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Two academic mini lectures</li> <li>➤ Planning conditions: Pre-task planning; limited planning</li> <li>➤ Planning time: one week for pre-task planning; three minutes for limited planning</li> </ul>	Complexity: <ul style="list-style-type: none"> <li>● Clauses/T unit</li> </ul> Accuracy: <ul style="list-style-type: none"> <li>● Errors/clause.</li> </ul> Discourse marking: <ul style="list-style-type: none"> <li>● Chaudron and Richards' (1986) discourse cues</li> </ul> Comprehensibility: <ul style="list-style-type: none"> <li>● Ratings of 25 undergraduates and 10 ESL specialists</li> </ul>	1. Both the number of discourse marking and the degree of explicitness of marking increased in planned production. 2. Significant differences were present in complexity. 3. No difference was found in accuracy. 4. Overt marking of discourse increased comprehensibility. 5. Complexity and accuracy of language could not explain differences in comprehensibility ratings.

			differences in comprehensibility.			
Ortega (1995b)	28 American learners of Spanish in the University of Hawaii /EFL	Less than 1 day (two oral tasks with a distracted task in between)	General hypotheses: 1. Planning will produce more syntactically and lexically complex language. 2. Planning will aid fluency of speech. 3. Planning will not increase accuracy of language.	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Two oral narrative tasks</li> <li>➤ Two planning conditions: no planning; pre-task planning</li> <li>➤ Planning time: 8 minutes</li> </ul>	<p>Syntactic complexity:</p> <ul style="list-style-type: none"> <li>● Words/utterance;</li> <li>● Propositions/utterance</li> </ul> <p>Lexical complexity:</p> <ul style="list-style-type: none"> <li>● Noun-verb ratio;</li> <li>● Lexical-to-grammatical words ratio;</li> <li>● Type-token ratio</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● TLU definite article;</li> <li>● TLU s-v agreement.</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Lexical dysfluency (partial words, repetitions, and self-corrections)</li> </ul>	<ol style="list-style-type: none"> <li>1. Significant increase in syntactic complexity was found in the planning condition.</li> <li>2. Lexical complexity measured by type-token ratio was significantly higher in the planned production.</li> <li>3. No increase in accuracy was found.</li> <li>4. The planned production had significantly lower rate of lexical dysfluency.</li> </ol>
Foster and Skehan (1996)	32 learners of English at a local college in the UK/ESL	3 weeks (one task per week)	General hypotheses: 1. Planning will have positive effects on performance in fluency, complexity and	<ul style="list-style-type: none"> <li>➤ Multifactorial design:</li> <li>➤ Three types of tasks: Personal information; Narrative; Decision making</li> <li>➤ Three planning conditions: Detailed</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Reformulation;</li> <li>● Replacement;</li> <li>● False start;</li> <li>● Repetitions;</li> <li>● Hesitations;</li> <li>● Pauses;</li> </ul>	<ol style="list-style-type: none"> <li>1. Strong effects of planning on fluency (for pauses and silence time but adverse effect on repetition and replacement) were found</li> <li>2. Clear effects on complexity (detailed</li> </ol>

			accuracy. 2. The effects of planning will be greater in detailed planning condition. 3. The effects will be greater for more cognitively challenging tasks.	planning; Undetailed planning; No planning. ➤ Planning time: 10 minutes	<ul style="list-style-type: none"> <li>● Silence in total</li> </ul> Complexity: <ul style="list-style-type: none"> <li>● Clauses/c-unit;</li> <li>● Variety of verb forms</li> </ul> Accuracy: <ul style="list-style-type: none"> <li>● Error-free clauses</li> <li>● Lexical errors</li> </ul>	planning >undetailed planning>no planning) were found. 3. There were significant differences between the no planning and undetailed planning condition in the personal and decision making tasks but not in the narrative task. 4. No difference was found between the two planning conditions.
Wendel (1997)	40 Japanese learners of English in a junior college in Japan/EFL	Not reported (each participant narrated one story and sat a cloze test as a proficiency test.)	General hypotheses: 1. Planning will lead to increased fluency, complexity, and accuracy in oral narratives.	<ul style="list-style-type: none"> <li>➤ Between-groups design</li> <li>➤ Two story-retelling tasks.</li> <li>➤ Two planning conditions: no planning; pre-task planning.</li> <li>➤ Planning time: 10 minutes.</li> </ul>	Syntactic complexity: <ul style="list-style-type: none"> <li>● Number of simple and complex T-units;</li> </ul> Lexical variety: <ul style="list-style-type: none"> <li>● Word families.</li> </ul> Accuracy: <ul style="list-style-type: none"> <li>● Percent of correct VP morphology</li> </ul> Fluency: <ul style="list-style-type: none"> <li>● Syllables/minute</li> <li>● Mean length of pause.</li> </ul>	1. An overall effect for planning was found. 2. Planning led to increased fluency. 3. Planning led to increased syntactic complexity but not increased lexical variety. 4. No increase in accuracy was found.

					Productivity: ● Number of main and subordinate clauses	
Wigglesworth (1997)	107 learners of English high and low proficiency levels tested in Australia/ESL	Less than 1 day (each participant did the whole set of the test)	What effects does planning have on discourse analytic ratings and measures for fluency, complexity, and accuracy?	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Tasks: A tape-mediated oral test with four sections of varying degree of difficulty. Planning time was provided in two sections and not in the other two.</li> <li>➤ Planning conditions: pre-task planning; no planning</li> <li>➤ Planning time: 1 minute</li> </ul>	Analytic ratings of fluency, grammar, and intelligibility. Complexity: ● Amount of subordination. Accuracy: ● Suppliance of plural –s; ● Verbal morphology; and ● Indefinite article. Fluency: ● Number of self-corrections. ● Type-token ratio	1. Planning led to higher fluency, complexity, and accuracy but the differences were only significant for the higher proficiency learners in the most difficult tasks. 2. No difference was found in analytic ratings.
Skehan and Foster (1997)	40 ELS learners at a local college in the UK/ESL	3 weeks (one task per week)	What effects does planning have on oral performance?	As in Foster (1996). <ul style="list-style-type: none"> <li>➤ Two variables investigated: planning time; post-task requirement</li> <li>➤ Planning time: 10 minutes</li> </ul>	Fluency: ● Number of pauses > 1 second. Complexity: ● Clauses/c-unit. Accuracy: ● Error-free clauses.	1. Planners were more fluent than non-planners. 2. Planners produced more complex language in the personal and decision-making tasks but not in the narrative task.

						<p>3. Planners were more accurate in the personal and narrative tasks but not in the decision task.</p> <p>3. There was some evidence that post-task led to greater accuracy in the decision-making task and greater fluency in the narrative task.</p>
Mehnert (1998)	31 learners of German at a university in London /EFL	2 weeks (each participant performed two tasks with one task per week)	Does more planning time result in improved oral performance?	<ul style="list-style-type: none"> <li>➤ Between-groups design</li> <li>➤ Two tasks: <ul style="list-style-type: none"> <li><input type="checkbox"/> Instruction (structured);</li> <li><input type="checkbox"/> Exposition (unstructured)</li> </ul> </li> <li>➤ Four planning conditions: <ul style="list-style-type: none"> <li><input type="checkbox"/> No planning;</li> <li><input type="checkbox"/> 1 minute planning;</li> <li><input type="checkbox"/> 5-minute planning;</li> <li><input type="checkbox"/> 10-minute planning</li> </ul> </li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of pauses</li> <li>● Total pausing time;</li> <li>● Length of run; and</li> <li>● Syllables per minute.</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Subordinate clauses/T-unit;</li> <li>● S-nodes/T-unit.</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses;</li> <li>● Errors/100 words.</li> <li>● Word order errors;</li> <li>● Lexical choice errors.</li> </ul> <p>Density of speech:</p>	<p>1. Planners were more fluent than non-planners. Generally, more planning time resulted in greater fluency.</p> <p>2. No significant differences were found either between planners and non-planners or among planners in complexity.</p> <p>3. Nonplanners produced significantly more errors than planners. More planning time did not result in fewer errors.</p> <p>4. Only 10 minute planning resulted in greater lexical density than non-planners in</p>

					<ul style="list-style-type: none"> <li>● Weighted lexical density.</li> </ul>	the structured task.
Ortega (1999)	64 learners of Spanish at the University of Hawaii/EFL	1 hour and 15 minutes (one familiarity task +two oral tasks with interviews in between)	<p>1. Does planning increase complexity, fluency, and accuracy of language?</p> <p>2. What do learners do when they plan?</p>	<ul style="list-style-type: none"> <li>➤ Repeated measures design.</li> <li>➤ Two oral narrative tasks.</li> <li>➤ Two planning conditions: no planning; pre-task planning</li> <li>➤ Planning time: 10 minutes</li> <li>➤ Retrospective interview</li> </ul>	<p>Complexity:</p> <ul style="list-style-type: none"> <li>● Words/utterance;</li> <li>● Type-token ratio;</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● TLU noun modifiers;</li> <li>● TLU articles;</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Pruned speech rate</li> </ul>	<p>1. Planned output had greater syntactic complexity and fluency.</p> <p>2. Hypothesized advantage for lexical range was not borne out.</p> <p>3. Results for accuracy were mixed. Increased accuracy for the use of noun-modifier agreement was found but not for articles.</p>
Foster and Skehan (1999)	66 ESL learners at a local college in the UK/ESL	Less than 1 day(each participant performed one task)	<p>General hypotheses: Different sources of planning (teacher-led; group-planning) and manipulation of foci of planning (content vs. language) have different effects on oral performance.</p>	<ul style="list-style-type: none"> <li>➤ 2x2 research design: source of planning; focus of planning.</li> <li>➤ Task: A decision-making task</li> <li>➤ Planning conditions: no planning; solitary planning; group planning; teacher-led planning</li> <li>➤ Planning time:10 minutes</li> </ul>	<p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/c-unit.</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of pauses and amount of silence;</li> <li>● Repetition,</li> <li>● False starts,</li> <li>● Reformulation and replacement</li> </ul>	<p>1. Teacher-led planning generated significant accuracy effects, while solitary planning had greater influence on complexity, fluency and turn length. Group based planning did not lead to performance significantly different from the control group.</p> <p>2. Different foci did not have significant effects on performance.</p>

Sangarun (2001)	40 Thai EFL learners at a high school in Thailand/EFL	1 day (for each participant there is one think-aloud training sessions, two oral tasks with one-hour intermission)	1. What are the effects of planning with different foci on the quality of speech 2. What are the planning processes resulted from the different foci?	<ul style="list-style-type: none"> <li>➤ Between-groups 4x2 design</li> <li>➤ Two tasks: Instruction task and argumentative task.</li> <li>➤ Four planning conditions: content focus, language focus, content and language focus, and no planning</li> <li>➤ Planning time: 10 minutes</li> </ul>	<p>Complexity:</p> <ul style="list-style-type: none"> <li>● S-nodes/T-unit;</li> <li>● Clauses/T-unit;</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses;</li> <li>● Errors/100 words.</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Speech rate A (unpruned);</li> <li>● Speech rate B (pruned speech);</li> <li>● Percentage of total pausing time.</li> </ul>	<p>1. Content focus and content and language focus planning promoted application of pre-task plans.</p> <p>2. Content focus and content and language focus conditions promoted complexity.</p> <p>3. The three planning conditions promoted accuracy in one of the two measures (i.e. percentage of error-free) and fluency (not in percentage of total pausing time), either in one or both tasks.</p>
Rutherford (2001)	31 ESL adult learners in New Zealand/ESL	Less than 1 day (each participant performed two tasks)	1. What effects does teacher-directed planning have on oral production? 2. What is the nature of planning?	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Two narrative tasks</li> <li>➤ Planning conditions: pre-task planning; no planning</li> <li>➤ Planning time: 5 minutes</li> </ul>	<p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/c-unit;</li> <li>● Length of c-unit.</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Number of error-free clauses;</li> <li>● Errors/100 words.</li> </ul>	<p>1. No statistically significant effect was found on either complexity or accuracy measures.</p>



Yuan and Ellis (2003)* <sup>20</sup>	42 Chinese EFL students at a Chinese university/EFL	Less than 1 day (a pre-test and an oral task)	What are the effects of pre-task planning and online planning on fluency, complexity, and accuracy in an oral narrative task?	<ul style="list-style-type: none"> <li>➤ Between-groups design</li> <li>➤ Task: An oral narrative task.</li> <li>➤ Planning conditions: no planning; pre-task planning; online planning</li> <li>➤ Planning time: 10 minutes for pre-task planning; unlimited for online planning.</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Syllables/minute;</li> <li>● Pruned syllables/minute.</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/T-unit;</li> <li>● Number of different verb forms; and</li> <li>● Mean segmental type-token ratio.</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Error-free clauses;</li> <li>● Percentage of correct verb forms.</li> </ul>	<ol style="list-style-type: none"> <li>1. The planning group was more fluent than the no-planning group (more pruned syllables per minute).</li> <li>2. The planning group produced more complex speech measured by clauses/T-unit (no difference on other two measures)</li> <li>3. There were no statistically significant differences found in accuracy between the pre-task planning and the no planning groups.</li> </ol>
Tajima (2003)	61 Korean learners of Japanese at a university in South Korea /EFL.	Less than 1 day (each participant performed a familiarization task, two experimental tasks, and filled out a questionnaire)	<ol style="list-style-type: none"> <li>1. What effects does provision of planning time have on oral performance?</li> <li>2. What role does attitudes towards planning play?</li> </ol>	<ul style="list-style-type: none"> <li>➤ Repeated measures design.</li> <li>➤ Two tasks: leaving a message on an answering machine.</li> <li>➤ Planning conditions: no planning; pre-task planning.</li> <li>➤ Maximum planning time: 10 minutes</li> </ul>	<p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses;</li> <li>● Target-like usage of particles</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/AS-unit;</li> <li>● Type-token ratio (TTR)</li> </ul> <p>Fluency:</p>	<ol style="list-style-type: none"> <li>1. There were positive effects of planning on fluency.</li> <li>2. There were positive effects on the general measure of accuracy (% of error-free clauses).</li> <li>3. There were positive effects on lexical complexity (TTR).</li> <li>4. Participants had higher scores on all the subjective</li> </ol>

<sup>20</sup>Studies marked with an asterisk investigated the effects of both pre-task planning and online planning. Only the results for pre-task planning are reported because of the limited space.

					<ul style="list-style-type: none"> <li>● Moras/second;</li> <li>● Pauses/minute;</li> <li>● Subjective rating on a 1-5 scale.</li> </ul> <p>Task completion:</p> <ul style="list-style-type: none"> <li>● Sum of scores on 3 subtasks;</li> <li>● Subjective rating on a 1-5 scale.</li> </ul>	<p>ratings in the planning condition.</p> <p>5. Syntactic complexity did not improve (negative effects of planning on clauses/T-unit).</p> <p>6. The positive feeling group produced significantly more accurate, more fluent language in the planned condition while the negative feeling group did not improve in any aspect in the planned condition.</p>
Kawauchi (2005)	39 Japanese learners of English at a university in Japan/ESL	Three weeks (one task per week)	<p>1. What are the effects of task repetition + planning on oral narratives of L2 learners?</p> <p>2. What role does proficiency level play in the effect of planning?</p>	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Three narrative tasks.</li> <li>➤ Three planning activities: writing a draft, rehearsal, and reading a model L2 input.</li> <li>➤ Planning time: 10 minutes</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Ratio of speech;</li> <li>● Repetitions.</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/T-unit;</li> <li>● T-unit length;</li> <li>● Number of subordinate clauses;</li> <li>● Number of word types.</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Use of past tense</li> </ul>	<p>1. Main effects of both planning and proficiency on fluency in both measures were found. Advanced learners made more repetitions in the planned condition than the unplanned condition.</p> <p>2. Main effects of both proficiency and planning on complexity in all measures were found. Planning did not</p>

					(copula be, regular verbs, irregular verbs)	aid the use of subordination for advanced learners. 3. Main effects of both proficiency and planning on accuracy were found. Advanced group did not get advantage from planning.
Skehan and Foster (2005)	61 ESL students at a college in London/ESL	Less than 1 day (each participant performed a task)	1. What are the effects of different forms of planning (i.e. guided vs. unguided planning) on oral performance? 2. Can the effects of strategic planning sustain? 3. Will introduction of surprise elements have an effect on performance?	<ul style="list-style-type: none"> <li>➤ Between-groups design</li> <li>➤ Task: a decision-making task.</li> <li>➤ Three planning conditions: no planning; guided planning; unguided planning.</li> <li>➤ Half of the three groups were given extra information after 5 minutes and the other half were not given extra information.</li> <li>➤ Planning time: 10 minutes</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● End of clauses pauses;</li> <li>● Mid-clause pauses;</li> <li>● Filled pauses;</li> <li>● Length of run;</li> <li>● Reformulations; and</li> <li>● False starts.</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Subordinate clauses/AS-unit;</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Error-free clauses and clause length.</li> </ul>	1. The guided planning condition led to more accurate speech than the no planning condition but only in the first 5 five minutes. 2. Guided planners produced more complex speech (subordinate clauses) than the other two groups in the first 5 minutes. 3. Both planning conditions produced more fluent speech (fewer end of clauses pauses) than the no planning group. 4. There is a significant reduction in performance levels after 5 minutes. 5. No clear influence of surprise information was

						found.
Elder and Iwashita (2005)	197 adult learners of English at a university in Australia/ESL	Less than 1 day (each participant completed one oral test)	What are the effects of planning time on student's performance of narrative tasks in a testing situation?	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Two narrative tasks in the Test of Spoken English.</li> <li>➤ A questionnaire asking about learners' perception of the difficulty of the task.</li> <li>➤ Planning conditions: no planning and pre-task planning</li> <li>➤ Planning time: 3.75 minutes</li> </ul>	<p>Assessors' ratings of fluency, complexity, and accuracy.</p> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of repetitions;</li> <li>● False starts;</li> <li>● Reformulations, and hesitations and pauses.</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/c-unit.</li> </ul>	<ol style="list-style-type: none"> <li>1. No effect for planning on assessors' ratings.</li> <li>2. No effect on accuracy, complexity, or fluency.</li> <li>3. Test-takers reported no difference in their perception of task difficulty in the planning and no planning conditions.</li> </ol>
Tavakoli and Skehan (2005)	80 adult EFL learners at an educational association in Iran /EFL	Not reported (each participant performed four tasks)	General hypotheses: 1. Degree of structure in a narrative will influence fluency and accuracy of performance but will not influence complexity.	<ul style="list-style-type: none"> <li>➤ 2x2x4 factorial design: Planning condition and language proficiency (two levels) were between-subject variables and task structure (four levels) was a within-subject variable.</li> <li>➤ Tasks: 4 narrative tasks</li> </ul>	<p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/AS-unit</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Speech rate;</li> <li>● Total silence;</li> <li>● Various repair measures; and</li> </ul>	<ol style="list-style-type: none"> <li>1. Planning had significant effects on fluency, complexity, and accuracy.</li> <li>2. More proficient learners' language was more fluent, complex, and accurate.</li> <li>3. Task structure had significant effects on complexity, accuracy, and fluency (number of pauses</li> </ol>

			<p>2. <i>Strategic planning will influence fluency, complexity, and accuracy.</i></p> <p>3. Proficiency level will not influence the effects of these variables.</p>	<p>(two structured and two unstructured)</p> <ul style="list-style-type: none"> <li>➤ Planning conditions: no planning and strategic planning.</li> <li>➤ Planning time: 5 minutes</li> </ul>	<ul style="list-style-type: none"> <li>● Length of run.</li> </ul>	and false starts).
Philp, Oliver & Mackey (2006)	42childESL learners in Australia /ESL	3 weeks (one task per week)	What is the relationship between pre-task planning and linguistic production in children's ESL classrooms?	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Three tasks: two-way information gap.</li> <li>➤ Three planning conditions: 0 minute, 2-minute, and 5- minute pre-task planning.</li> </ul>	<p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of target-like c-units.</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Amount of subordination and coordination;</li> <li>● Percentage of words as lexical verbs.</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of reformulations and false starts/turn.</li> </ul>	<p>1. No effect was found for fluency.</p> <p>2. 5-min planning resulted in more grammatically complex language than the other two conditions. No effect was found on lexical complexity.</p> <p>3. No significant effect was found for accuracy.</p>
Gilabert (2007)	48 EFL learners at a university	4 days (2 sessions with two stories in	How does increasing the cognitive	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Two oral narratives</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Pruned speech rate.</li> </ul> <p>Complexity:</p>	<p>1. Planning produced higher speech rate for both tasks.</p> <p>2. Significant effects for</p>

	inBarcelona/EF L	one session)	complexity of tasks along planning time and the Here-and-Now variable affect fluency, complexity, and accuracy of production?	<ul style="list-style-type: none"> <li>➤ Four conditions:</li> <li><input type="checkbox"/> Planned Here-and-Now;</li> <li><input type="checkbox"/> Unplanned Here-and-Now;</li> <li><input type="checkbox"/> Planned There-and-Then;</li> <li><input type="checkbox"/> Unplanned There-and-Then;</li> <li>➤ Planning time: 10 minutes</li> </ul>	<ul style="list-style-type: none"> <li>● Lexical richness</li> <li>● S-nodes/T-units</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of self-repairs.</li> </ul>	<p>lexical richness were also found for both tasks in the planned condition.</p> <p>3. No significant effect on syntactic complexity was found in the planned condition.</p> <p>4. Planned condition produced less self-repair but the difference was not significant.</p> <p>5. Learners were significantly more fluent in the Here-and-Now task.</p> <p>6. Increasing task complexity along the +/- Here-and-Now variable reduced lexical complexity and did not affect syntactic complexity. It had a positive effect on accuracy.</p>
Guara-Tavares (2008)	50 EFL learners at a university in Brazil/EFL	Not reported (one session for proficiency test, a speaking span	<p>1. What are the effects of pre-task planning on L2 speech performance?</p> <p>2. What role does</p>	<ul style="list-style-type: none"> <li>➤ Between-groups design</li> <li>➤ Task: two narrative tasks.</li> <li>➤ Planning conditions: no planning and pre-task planning</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Speech rate (pruned and unpruned);</li> <li>● Pauses/c-unit;</li> <li>● Total percentage of pausing time.</li> </ul>	<p>1. Pre-task planning had impact on L2 accuracy and complexity, but not on fluency.</p> <p>2. Under the no planning condition, working memory</p>

		test, one session for the planning effect)	working memory play in planned and unplanned performance and the process of planning?	<ul style="list-style-type: none"> <li>➤ Planning time: 10 minutes</li> <li>➤ Retrospective interview (for all).</li> <li>➤ Retrospective online protocol (for experimental group).</li> </ul>	<p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Errors/100 words;</li> <li>● Percentage of error-free clauses.</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Clauses/c-unit</li> </ul>	significantly correlates with L2 speech fluency (for the experimental group), L2 accuracy (for the control group). 3. Under planning condition, working memory correlates with L2 fluency and complexity. 4. Higher span learners used significantly more metacognitive strategies than lower span learners during planning time.
Mochizuki and Ortega (2008)	56 beginning level students of English at a high school in Japan/EFL	Less than 1 day (participants performed the task in pairs)	What effect does teacher-led attention to a specific grammatical feature (relative clauses) have on performance of this feature in an oral task?	<ul style="list-style-type: none"> <li>➤ Between-groups design</li> <li>➤ Task: a one way story-retelling task.</li> <li>➤ Three planning conditions: no planning; guided planning; unguided planning</li> <li>➤ Planning time: 5 minutes.</li> <li>➤ Students received handout about how to construct relative</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Mean number of words/minute.</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Mean length of T-unit;</li> <li>● Mean number of clauses/T-unit; and</li> <li>● Number of relative clauses/T-unit.</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Amount and quality</li> </ul>	<ol style="list-style-type: none"> <li>1. Trend for non-guided planners to be more fluent than the other two groups was found.</li> <li>2. No group difference in complexity was found.</li> <li>3. Guided planners produced more and more accurate relative clauses than the other two groups.</li> </ol>

				clauses.	of relative clause use.	
Wang (2009)	77 Chinese learners of English at a university in Hong Kong/EFL	Less than 1 day (each participant completed a pre-test, two oral tasks, and an interview)	What are the effects of 5 types of planning and two task structures on L2 speaking performance?	<ul style="list-style-type: none"> <li>➤ Between groups design for planning; Repeated measures design for tasks</li> <li>➤ Two types of tasks: Structured narrative; unstructured narrative</li> <li>➤ Five planning conditions: watched<sup>21</sup>; watched+strategic; task repetition<sup>22</sup>; online planning<sup>23</sup>; watched online planning.</li> <li>➤ Planning time:3 minutes</li> </ul>	<p>Complexity:</p> <ul style="list-style-type: none"> <li>● Pruned total words;</li> <li>● Pruned morphemes/ AS-unit;</li> <li>● Total number of subordination clauses; and</li> <li>● Verb infinitives divided by total AS-unit</li> </ul> <p>Lexical diversity: D</p> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Number of error-free clauses;</li> <li>● Ratio of error-free clauses</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Speech-rate: pruned words per minute.</li> <li>● AS-end-pause;</li> <li>● Reformulation</li> </ul>	<ol style="list-style-type: none"> <li>1. Compared with the control group the watched + strategic planning condition had a medium to large positive effect on complexity .</li> <li>2. It had a large positive effect on the quantity part of accuracy measures (error-free clauses).</li> <li>3. It had a near-large positive effect on speech-rate.</li> <li>4. No effect was found for lexical diversity (D) and other fluency measures.</li> <li>5. The unstructured task led to greater complexity and fluency measured by speech rate. Task structure did not have significant effects on accuracy.</li> </ol>

<sup>21</sup>The speaker watched the video once before narrating the story.

<sup>22</sup>The speaker narrated the story twice.

<sup>23</sup>The speaker narrated the story while watching the video that had been edited to play at a slower rate.



Mehrang and Rahimpour (2010)	62 learners of English at a language institute in Iran/EFL	Less than 1 day (each participants performed two tasks)	What is the impact of planning conditions on the oral performance of EFL learners while performing structured vs. unstructured tasks?	<ul style="list-style-type: none"> <li>➤ 2x2 factorial design: Planning was a between-subject design. Task structure was a within-subject design.</li> <li>➤ Two narrative tasks: one structured and one unstructured.</li> <li>➤ Planning conditions: no planning and pre-task planning.</li> <li>➤ Planning time: 5 minutes</li> </ul>	<p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Ratio of error-free T-units</li> </ul> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Words/ minute</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Number of lexical or 'open class' words in a text/total words</li> </ul>	<ol style="list-style-type: none"> <li>1. Planning had no effect on accuracy or fluency, but led to more complex performance in the unstructured task.</li> <li>2. Task structure did not affect accuracy and complexity while promoting fluency under the planned condition.</li> </ol>
Wigglesworth and Elder (2010)	90 ESL learner in a testing situation in Australia/ESL	1 day (each participant completed an IELTS oral test)	Does different amount of planning time cause difference in quality of oral discourse in a testing situation?	<ul style="list-style-type: none"> <li>➤ Repeated measures design</li> <li>➤ Task: three tasks from IELTS Part 2<sup>24</sup></li> <li>➤ Proficiency: intermediate and advanced determined by IELTS score.</li> <li>➤ Planning conditions: no planning; one-minute planning; two-minute planning.</li> </ul>	<p>IELTS ratings</p> <p>Fluency:</p> <ul style="list-style-type: none"> <li>● Fluent versus disfluent speech</li> <li>● Filled and unfilled pauses</li> <li>● Self repairs</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Error-free AS units</li> <li>● Error-free clauses</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Proportion of</li> </ul>	<ol style="list-style-type: none"> <li>1. There was no statistically significant effect for either task, planning time, or proficiency in global ratings.</li> <li>2. There was no significant effect of planning found in fluency.</li> <li>3. No significant effect of planning was found in complexity.</li> <li>4. No significant effect of planning was found in</li> </ol>

<sup>24</sup>In IELTS Part 2, a test taker is given 1 or 2 minutes to prepare and then talks on a given topic.

					<p>dependent clauses per AS unit</p> <ul style="list-style-type: none"> <li>● Percentage of subordinate clauses to AS units</li> </ul>	accuracy.
Nakakubo (2011)*	147 learners of Japanese at university-level institutions in America and Japan/EFL and ESL	Not reported(each participant completed a demographic questionnaire, a proficiency test, two narrative tasks, an interview)	<p>1. What are the effects of pre-task and online planning on L2 oral narrative task performance?</p> <p>2. What are the effects of time pressure on L2 oral production?</p>	<ul style="list-style-type: none"> <li>➤ Between-group design</li> <li>➤ Two oral narrative tasks</li> <li>➤ Planning conditions: no planning +/-time pressure, pre-task planning +/-time pressure, online planning, and pre-task planning +online planning.</li> <li>➤ Planning time: 10 minutes for pre-task planning; unlimited time for online planning.</li> </ul>	<p>Fluency:</p> <ul style="list-style-type: none"> <li>● Number of moras per minute</li> </ul> <p>Complexity:</p> <ul style="list-style-type: none"> <li>● Number of clauses per T-unit;</li> <li>● Type-token ratio;</li> </ul> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>● Percentage of error-free clauses (global)</li> <li>● Percentage of correct use of particles (specific)</li> </ul>	<p>1. Planning did not impact fluency significantly.</p> <p>2. No significant difference was found in syntactical complexity. In the no planning condition learners produced significantly more lexically complicated narrative than in the planning condition.</p> <p>3. No significant difference was found in accuracy.</p> <p>4. Time pressure had a significant effect. Under time pressure learners spoke faster, but there was no significant effect on complexity or accuracy.</p>



