Metacognitive Strategy Instruction and Pre-task Planning: Impact on L2 Argumentative Writing Ability

Sara Amani

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
UNIVERSITY OF AUCKLAND, FEBRUARY 2014

ABSTRACT

Research on language learning strategy instruction has documented the effectiveness of helping less successful language learners improve their performance through strategy instruction (e.g. Oxford, 1990; Wenden, 2002). Fewer studies, however, have investigated the impact of an explicit instructional training of strategies on conducting written tasks. For this reason, more research is required into the investigation of a relationship between writing strategy use and second language (L2) writing ability, given the crucial role played by the latter in the academic success and, by extension, the educational and career prospects of L2 learners. The current study was conducted to explore the possibility of designing a preparatory writing metacognitive training program which would incorporate writing strategy training activities into everyday classroom language instruction to introduce learners to writing-specific strategies. In this way, it was hoped to maximise L2 learners' argumentative writing performance by raising their metacognitive awareness and providing them with sufficient time to plan content and form.

Two studies were conducted: first in a university in Australia (n = 35), and then in a language school in Iran (n = 70). Contrary to the typical standard designs in extant task-based planning research, which requires the participants in experimental and control groups to perform one single task under different planning conditions, the design of this quasi-experimental study sought to address development by the inclusion of a four-session unit of instruction which was specifically aimed at comparing the participants' performance with and without metacognitive strategy instruction. Using mixed methods, the quantitative findings obtained from writing essays were triangulated with the quantitative technique of Likert-scale questionnaires and the qualitative in-depth interpretations of think-aloud protocols.

The results provided evidence of the effectiveness of metacognitive training on the lexical complexity of written products on the one hand and the general writing proficiency on the other. Furthermore, the complementary data obtained from think-aloud protocols and questionnaires contributed to our understanding of what actions learners engage in when writing an essay and how metacognitive instruction can impact their conscious efforts to

produce a text which is grammatically correct, logically coherent and situationally appropriate. The resulting contribution will be twofold: Firstly, writing instruction will embed a preparatory metacognitive training program in the content matter to ensure connectivity. Secondly, it will justify that introducing learners to writing task-specific metacognitive strategies is an important step forward to resolve major executive and practical problems in essay writing courses at secondary and tertiary levels.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to all those people who have given me such affirmative support and assistance over the course of these past few years, as it is certain that without their help I would not have been able to satisfactorily complete this study.

First and foremost, I am privileged to express my deepest gratitude to my main supervisor, Dr Jenefer Philp for her invaluable comments and suggestions, along with her informed advice and expertise in the writing of this thesis. I especially thank her for her cooperation and continuing support from the earliest stages of gestation to the final stages of refinement. I also remain greatly indebted to my co-supervisor, Dr Rosemary Wette for her unfailing efficiency, her decisive clarity and the discernment of her creative guidance. Her astute eye, her immaculate attention to detail and her insightful comments have helped enormously to polish both the format and the academic quality of this thesis. Last but not least, I must also thank Dr. Rebecca Adams for the active assistance she gave me as my first co-supervisor over the initial two years of my thesis. Her academic strength and support allowed me to gain the essential ground needed, and to enable me to more easily proceed to this conclusive point in my studies.

On a more personal note, my most sincere thanks go to my friend, Nebojsa Gosic, for his forbearance, encouragement and support. My deepest gratitude then goes to another very dear friend of mine, Anne Jones Watson, whose inspirational love and support have always given me the courage to continue with the task of finishing my thesis, regardless of the obstacles which appeared at times to tower before me. In regard to those individuals who placed these obstacles in my path, as I approached the last and most important year of this thesis, I would like to wholeheartedly thank them for challenging my equilibrium at such a vitally important time in my studies. They have inadvertently helped me to harden my resolve and to ensure my own personal success. Their actions have shown me that I am someone who cannot be deflected from my focus towards excellence. Finally and most importantly, my truly heartfelt thanks go to my precious family who has always been the driving force behind my perseverance and sense of excitement as I have progressed through all my academic achievements and endeavours.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

Accuracy: The extent to which the language produced in performing a task conforms with target language norms. The measure of accuracy will be based on the percentage of accurately used verbs and clauses that do not contain any syntactic, morphological and lexical choice errors (Ellis, 2003).

Complexity: The extent to which the language produced in performing a task is elaborated and varied (Ellis 2003, p. 340). The measure of complexity will be based on syntactic complexity, i.e. the ratio of clauses to T-units in the participant's production, syntactic variety, i.e. the total number of grammatical verb forms (tense, modality, voice) used in the task, and the type-token ratio of each segment in the participants' narratives.

Metacognitive strategy use: It refers to thinking about or knowledge of the learning process, planning for learning, monitoring learning while it is taking place, or self-evaluation of learning after the task has been completed (O'Malley & Chamot 1990, p. 231).

No Planning: It refers to the learner's immediate performance of the task without having enough time to plan for the task.

Pre-task planning: It refers to the process by which learners plan what they are going to write before commencing a task. Pre-task planning can attend to propositional content, to the organization of information, or to the choice of language. It contrasts with the on-line planning (Ellis 2003, p. 348).

Second language writing (L2 writing): Any writing done in a language other than the writer's native language. Most second language writing research refers to writing in English as a second or foreign language (Reichelt, 1999).

Strategy: The mental and communicative processes that learners categorize in performing a task at hand (Nunan, 2004).

Task: A task is a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed (Ellis 2003, p. 16).

T-unit (also Minimal Terminable Unit): A measure of the linguistic complexity of sentences, defined as the shortest unit (the Terminable Unit, Minimal Terminable Unit, or T-Unit) which a sentence can be reduced to and consists of one independent clause together with whatever dependent clauses are attached to it. For example, the sentence *After she had eaten, Kim went to bed* would be described as containing one T-Unit (Richards et al., 1985, p. 211).

Working memory: A limited resource that is drawn on for both storing information and carrying out cognitive processes which still remain nonautomatised. It includes managing tasks that require problem solving and decision making by making use of phonological representations, visual or spatial representation, and semantic memory (Hayes, 1996).

Writing process: Writing is the production of thought for oneself or others under the direction of one's goal-oriented metacognitive monitoring and control, and the translation of that thought into an external symbolic representation (Hacker, Keener & Kircher, 2009).

CHAPTER 1. INTRODUCTION

This chapter provides information about cognition and metacognition in second language writing processes, as they are central to the theoretical framework which underpins this study. It outlines the context of English for academic purposes (EAP) provision in Australia and Iran, the two contexts in which this research was conducted. A rationale and justification for the study are provided. An explanation of the general research approach is given and the chapter ends with an overview of general structure of the thesis.

1.1. Cognitive and social theories of learning

Approaches to teaching writing have dramatically changed over the past sixty years. Product-centred pedagogy, which Young (1978) termed "current-traditional", was the main approach in the teaching of writing throughout the 1950s and 1960s. It emphasised the learning of particular modes of discourse, and their application in the composition of five-paragraph essays. There was a strong focus on usage (syntax, spelling, punctuation) and style (economy, clarity, emphasis). In this teacher-oriented pedagogy, teachers graded completed compositions, but students were not provided with the opportunity to receive feedback or revise. However, in the 1970s, as part of the growing influence of cognitive and expressive theories of learning, process approaches to pedagogy in second language writing replaced product-oriented pedagogies as the dominant approach (Emig, 1971; D. M. Murray, 1972). In these approaches, the focus of interest shifted to the inclusion of a variety of writing tasks, the improvement of writing assessment, the role of teachers as facilitators and students as explorers of their own "voice". For the first time, students were allowed to choose their own topics and were encouraged to revise their writing (Matsuda, 2003a). The writing process research was formally introduced into second language writing studies by Zamel (1976). Subsequently, process approaches attained and maintained a dominant status in the field of second language (L2) writing research and pedagogy (B. Kroll, 1978; Raimes, 1983, 1985).

By the late 1980s, influenced by advances in psychology, linguistics and general education, social theories began to influence composition studies for first language (L1) writers and

subsequently second language writing (e.g. Berlin, 1988; Bizzell, 1986; Faigley, 1986). Although the importance of social interaction and the communicative purpose of texts began to be recognised in this "sociocultural turn", its advocates did not argue against process pedagogy, but rather they reinterpreted and broadened the conception of writing. The recognition of social and cultural aspects of composing processes had a bearing on register studies, discourse studies, genre studies, and corpus studies. In Australia, for instance, post-process, post-cognitivist theories led to genre-based pedagogies, consistent with Halliday's Systemic Functional Linguistics (1978), which began to feature in writing across the curriculum instruction to prepare students for their future literacy encounters.

Around the same time, studies of oral and written L2 performance within task-based forms of instruction attracted considerable attention, with their primary concern for the effects of task design and implementation variables on the fluency, complexity and accuracy of language (e.g. Crookes, 1989; Ellis, 1987). Drawing on cognitive aspects of learning, task-based instruction supports the contention that L2 performance is characterised by competing demands, and that increasing attention to one aspect of language is to the detriment of others (Skehan, 1998). This dominance of the computational metaphor in task-based research sees interaction as a source of input to the learner which creates a context for acquisition (Ellis, 2003). An alternative paradigm for viewing tasks is sociocultural (Lantolf, 2000) in which participation is more central than acquisition and is researched via microgenetic analysis, i.e. a detailed analysis of how new forms arise out of interpersonal activity while learners are performing a task. In this light, learning takes place in social events, linguistic forms and functions are internalised with time and practice, and development subsequently happens out of social interactions with a progression from the inter-mental to the intra-mental.

1.1.1. Cognition and metacognition in writing processes

Contemporary writing models no longer view writing as a linear and simplistic activity. Instead they recognise it more as linguistic, cognitive, affective, behavioural, and social (Graham, 2006; McCutchen, 2006; Prior, 2006). With the emergence of metacognitive theories in learning in the 1970s (e.g. Flavell, 1971, 1976), our understanding of complex cognitive processes involved in writing has accordingly evolved. The definitions of metacognition initiated by Flavell (1976) and Flavell and Wellman (1977) are also

generally followed in writing research. These definitions include metacognitive knowledge (i.e. knowledge about persons, tasks, strategies), and metacognitive monitoring and self-regulation (i.e. planning, monitoring, evaluating) (Flavell, Miller, & Miller, 2002). Asserting that "writing is applied metacognition", Hacker, Keener, and Kircher (2009) propose a metacognitive theory of writing from which their definition of writing is derived. They define writing as "the production of thought for oneself or others under the direction of one's goal-oriented metacognitive monitoring and control, and the translation of that thought into an external symbolic representation" (p. 154).

Since Devine's (1993), and Devine, Railey and Boshoff's (1993) first attempts to highlight the role of metacognition in both L1 and L2 writing, a good number of studies have focused on the implications of metacognition and cognitive models in writing task performance. According to Flower (1994), strategy instruction is metacognitive only if it actively engages students in understanding, monitoring and controlling their own learning. To date, a number of educational interventions and writing curricula have been designed which cue appropriate metacognitive strategy use related to planning, drafting, revising, and editing. Two extended examples in L1 writing are Cognitive Strategy Instruction in Writing (Englert, Raphael, Anderson, Anthony, & Stevens, 1991) and instruction in task-specific procedures (Smagorinsky, 1991). The former emphasises a combination of cognitive instruction and metacognitive learning and underlies the role of dialogues in writing development, direct explanation of writing strategies and modelling them, scaffolded instruction, and procedural facilitation. The latter suggests combined instruction of declarative knowledge of form plus task-specific procedural knowledge related to content plus procedural knowledge related to form. In a similar vein, the research on strategy instruction in L2 writing has largely supported metacognitive teaching to be an integral part of writing instruction (e.g. He, 2005; Mayer, 2001; Xiao, 2007).

From a metacognitive perspective, writing is a highly complex enterprise, one which requires the knowledge and regulation of many cognitive activities. Garner (1987) made a distinction between cognition and metacognition, in which cognitive skills are necessary to perform a task, while metacognition is necessary to understand how the task is performed. In exploring the relationship between metacognition and cognition, N. J. Anderson (2007) has asserted that cognitive processes often operate at an unconscious and automatic level and are used for the manipulation or transformation of the language and

task, whereas metacognitive skills require a conscious processing and are used to provide executive control over the execution of a task. Motivated by the distinction between cognition and metacognition, the current study investigates the concept of explicitness in metacognitive strategy instruction to help L2 learners to move to an upper level of thinking about thinking, to learn how to learn writing, and to reach a consciousness of what is happening as they perform the written task. It aims to apply metacognitive instruction to provide an alternative solution to the problems L2 learners face when writing, particularly academic essays, and become knowledgeable about their cognitions.

1.1.2. Planning in writing

The theoretical perspective in this study was mainly informed by the cognitive theories of writing (Hayes, 1996), which focus on the role of working memory in writing and the interplay of planning, text production, and revision. The literature is replete with studies exploring L2 planning processes as part of task-based research. The studies investigating the impact of planning on language production draw on information processing models which claim that there are limits on the amount of information that human beings can process, and thus language learners tend to prioritize one aspect of language over another (Newell & Simon, 1972; Robinson, 1995). The cognitive models of task complexity in task-based performance (e.g. Skehan & Foster, 2001) explain that the provision of planning, whether of the pre-task or unpressured within-task kind, eases the burden on the limited working memory capacity. In this way, learners are given the opportunity to engage in controlled processing and to process multiple systems, and as a result the quality of their output is enhanced.

Task planning is an implementation variable which has yielded relatively consistent effects on L2 writing production. A number of studies (e.g. Ellis & Yuan, 2004; Ojima, 2006; Ong & Zhang, 2010) have shown that when learners have the opportunity to plan a task prior to task performance, the writing they produce is more complex than the no planning condition, but not more accurate. Other studies (e.g. Ellis & Yuan, 2005; Khomeijani Farahani & Meraji, 2011) have reported that the writers' moment-by-moment online planning in the process of formulation and monitoring as they perform the task has predictable effects, particularly on the accuracy of written products.

Almost all these studies investigating the impact of planning on writing are typically concerned with how written production (rather than language acquisition) is enhanced. The standard design used in such task planning research involves experimental and control groups which perform tasks under different planning conditions. However, as Ellis (2005) has argued, these studies "cannot address acquisition" (p. 28). The effects of planning on acquisition can only be investigated by employing a design which incorporates a treatment or instruction, as a result of which different kinds of evidence of change may be pinpointed. To the best of my knowledge, the only study which has made use of such a design in writing tasks is Jong (2009) which assesses the effects of task type and task repetition on Korean EFL learners' writing performance in terms of fluency, accuracy, and lexical complexity. The results in Jong's study showed strong effects of task repetition on both the products and processes of learners' writing. Inspired by the second type of enquiry into the effects of planning on acquisitional changes in L2 writers' knowledge representation, the present study addresses the extent to which instruction can help learners to sustain the effects of planning on performance over an extended period of time.

1.2. Academic writing instruction in ESL and EFL contexts

Kroll (2003) has argued that English-dominant cultures are progressively driven by written and digital literacy, and thus "the pursuit of English entails a pursuit of written English" (p. 1). In higher education institutions, students' writing ability is generally considered as having a strong influence on their academic and professional success. The volume and variety of writing tasks required in undergraduate general education and disciplinary courses places additional pressures on L2 learners who are required to perform their writing in response to curricular demands. Meeting these demands is at times exacerbated by the differences in writing qualities valued by different disciplines which have varied discourses of academic inquiry. These diverse standards are likely to cause further hardship for L2 writers who may have been successful in handling the literacy demands of academic studies in their home countries, but who find the demands of study in English medium universities challenging.

Furthermore, foreign language or second language contexts per se have been considered a shaping factor on writing approaches and L2 learners' written performances (Manchón, 2009; Ortega, 2009). A number of different constraints and resources have been shown to

be potentially influential. Analysis of micro- and macro-context influences show that the attitudes of the community to learning writing in English, the role of writing English in the wider community, the instructional methodology used in the classroom, and practice opportunities characterise L2 writing development differently. In a typical FL writing context, since learners do not have ready access to the target language outside language classrooms, the FL writing may be less purposeful and less needs-driven. In an SL setting, by contrast, learners have less difficulty in seeing the relevance of writing English to their immediate use of the language beyond the confines of classrooms. Therefore, they are expected to exhibit comparatively higher levels of writing competencies. As a result, there is a need for research studies which empirically investigate various aspects of L2 writing ability across writing contexts. Because the participants in the present study are from various language backgrounds and nationalities that have travelled to Australia to pursue their tertiary education, i.e. SL context, or Iranians learning academic English in their home country, i.e. FL context, a background to these two contexts is presented.

1.2.1. Australia

Since English is currently the main lingua franca in international research and scholarship in many academic disciplines, more than 1.2 million currently study in English-speaking countries in English, which now "seems to [be] becoming less a language than a basic academic skill for many users around the world" (Hyland, 2013, p. 54). In 2010, international students represented 21.5% of higher education enrolment in Australia (Roberts, 2012). This places Australia as the third most popular destination for tertiary education (7%) after the United States (17%) and the United Kingdom (13%). The student population in Australia is increasingly diverse, but currently the majority come from China and India.

Some international students from non-English medium settings are required to meet the requirements of IELTS or TOEFL scores which are specified by the particular university they intend to seek admission to, with most universities requiring an overall IELTS score of 6.0 for entry to undergraduate studies. A number of disciplines have higher requirements (e.g. education, journalism, interpreting). The equivalent score on paper-based TOEFL test is a minimum test score of 580 with a Test of Written English (TWE) score of at least 4.5. The equivalent internet-based TOEFL test is a minimum test score of

90 overall, at least 22 in writing and no section lower than 20. Other recognised entry pathways include successful completion of a recognised English language bridging program at a particular level of achievement. Students can also gain admission through the Australian secondary school system by meeting the academic literacy requirements which are in place for all students.

Instruction in the primary and secondary schools and in the EAP bridging programs commonly uses genre-based approaches which are linked to the basic principles of systematic functional linguistics. In this type of instruction, teachers take students through the initial step of understanding the context for writing, where their learning is scaffold by direct instruction as well as a variety of exercises including role plays, jigsaws, and vocabulary building exercises. The second step involves assisting students to model and deconstruct the text through the use of various text models of the chosen genre. In the next step, students construct texts collaboratively with their peers or their teacher. Finally, they independently produce their own texts (Johns, 2003).

1.2.2. Iran

English as a Foreign Language (EFL) is introduced within the Rahnamaei (literally meaning guidance or orientation) school system, which is an equivalent of junior high school or middle school. This is a period of three years which covers grades six to eight, for students aged 11 to 13. Similar curriculum standards are followed in the three years of high school and one year of pre-university program. At the tertiary level, if English is not studied as a separate major in three branches of English Literature, Translation, and Teaching English as a Foreign Language (TEFL), it then primarily consists of two parts for all students in any major: (1) a general English course of three-unit credit, and (2) an English for Specific Purposes (ESP) course of four-unit credit, such as English for Engineering, English for Medical Sciences, or English for the Social Sciences. The first course is aimed at strengthening students' general reading comprehension of written materials in their majors, while the second course is targeted at familiarising students with English specific concepts and technical terms related to their specific fields.

English language teaching in Iran is restricted to "classroom practices, including the use of particular textbooks and the teacher's management of classroom work, without substantial

support from social contexts outside the classroom" (Ghorbani, 2009, p. 132). Opportunities for real-life communications outside classroom situation are rare. As Hayati and Mashhadi (2010, p. 34) assert, "language practices are often based on language *usage* rather than language *use*" (emphasis in original). Because of this emphasis on sentence-level grammar, vocabulary and reading skills, students who want to develop skills in oral and written communication often take English courses in private language schools, which are self-funding but government-affiliated. At present, there are more than 5000 foreign language schools in the country, 200 of which are situated in Tehran (Ala Amjadi, 2012).

Borjian (2013) maintains that the existing system of English education in post-revolutionary Iran (after 1979), is marked by two diverging forms of English: (1) an indigenised model, and (2) an internationalised or Anglo-Americanised model. The former is used by state-run education programs, whereas the latter is used by the private-run education programs. Each of these employs a different kind of curriculum, pedagogy and type of course materials. Within the indigenised model, whether in public or private (termed as non-profit) schools or universities, home-grown structure-oriented textbooks which are complied, developed, authorised, and published by a state-run publishing house within the Ministry of Education are the only source of language input. Here, structural features like phonology, morphology, and syntax are emphasised. Among the language skills, reading ability is primarily fostered and given a higher priority, due to the main pre-defined goal that the High Council of Cultural Revolution (HCCR) set for the English education in Iran. This goal has been defined as "enabling students to read technical texts in universities" (High Council of Cultural Revolution 2002, cited in Atai & Mazlum, 2012, p. 10). The rationale for this goal is to benefit from the latest scientific knowledge and technological information located in written materials and to gain industrial, economic, and agricultural self-sufficiency (p. 15). Although the textbooks devote some sections to writing, the writing exercises are merely practices in grammatical features, rather than using the language in some functional ways in real-life situations, such as writing a diary, writing a letter to a friend, or writing argumentative essays. Oral and aural skills and the cultural pragmatic knowledge of the language are not emphasised or addressed in the school final exams, and the University Entrance Exam (Konkoor), which is a high-stakes multiple choice test. The approaches to language teaching pedagogy are also traditional, and utilise translation, repetition, memorisation, recitation, and reproduction (Ghorbani, 2009; Hosseini, 2007).

Within the internationalised model which is used in private language schools, published textbooks from USA and UK are commonly used, with 'Headway', and 'Interchange Series' two of the most popular. These books are based on communicative principles of language teaching and learning, and thus the educational goal is to help students to use English for functional purposes using a large variety of situationally or interactionally authentic communicative tasks. In this way, these privately-run language institutions supplement the curriculum offered at state-run schools.

In short, the main emphasis in the educational system of Iran, as stated by Riazi and Mosalanejad (2010), is primarily on "acquiring knowledge in the form of rote learning and memorisation, rather than constructing it through higher-levels of cognitive skills such as analysis and synthesis" (p. 23). Such being the case, writing activities which could vary from controlled writing to free writing are neglected and are only limited to a number of isolated and decontextualised grammatical exercises or responses to reading comprehension questions. When they enter the collegial-level institutions, the students who have learned English within this seven years of schooling are not generally able to communicate either orally or in written form, although they have successfully passed the high-stakes University Entrance Exam. Iran is not isolated from ongoing advances in the digital revolution and an ever-increasing need of communication with the world beyond its borders for business, research, and social purposes. As a result, and because of the currently dominant position of English, there is a need for students who wish to actively take part in disciplinary and other international communities to be able to function adequately in English across all four skills as well as other components of communicative competence, and for their abilities to go beyond the mastery of reading technical texts and translation.

1.3. Aims and rationale

Studies of task-based language teaching, particularly the role of planning in second language learning, have been mainly conducted with regard to oral rather than written production (Foster & Skehan, 1996, 1999; Mehnert, 1998; Ortega, 1999; Skehan & Foster, 1997, 1999; Yuan & Ellis, 2003). The current study, however, aims to contribute to knowledge in this area by both examining the features of texts and exploring the actions of

writers, i.e. a combination of a text-oriented and a writer-oriented approach. The primary purpose of the study is to examine the impact of metacognitive strategy instruction and pre-task planning on L2 learners' written performance; that is, it seeks to explore whether an enhanced writing performance can be achieved through the implementation of strategy-based intervention. In so doing, four experimental groups in the first study and five groups in the second study (four experimental and one control group) were used to examine the effects of metacognitive strategy instruction versus rhetorical modes of discourse instruction on writing performance. In taking on the cognitive orientation as the theoretical framework of the study, participants' written data were assessed in order to examine how their textual writing practices were constructed under different instructions. One of the main hypotheses in this study is that second language writing ability is centrally linked with the appropriate writing strategies. Studies with similar research aims have involved the construction and validation of writing strategy questionnaires (Purpura, 1997, 1999; Tsai, 2004); however, this study investigates the relationship between metacognitive strategy use and L2 academic writing ability utilising detailed and holistic rating rubrics, and models the use of these strategies using instruments rooted in theories of cognition.

The study also aims to investigate the actions that participants engage in while they are producing a text, and how metacognitive strategies operate for participants who have received the instruction compared to those who have not. Thus, the analysis takes into consideration the interplay of factors in both the process and product of writing. To pursue these two aims, a mixed-method methodology (Creswell, 2009; Teddlie & Tashakkori, 2009) was used which was primarily quantitative in its approach and partially qualitative. In order to collect the required data, multiple methods including participants' written texts, questionnaires, and think-aloud protocols were used. Quantitative results from the questionnaires and the qualitative findings from the think-aloud verbalisations were triangulated with the quantitative analysis of the written texts to help interpret the overall findings of the study. Although a small number of studies have investigated the relationship between cognitive and metacognitive strategies used by L2 writers and their performance on L2 writing task performance (e.g. N. J. Anderson, 2007; Y. Zhang, 2010), the present study is among the first to investigate metacognitive strategies-based instruction for academic writing in both EFL and ESL contexts.

1.4. Research Questions

In this study, the following research questions will be addressed:

1. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *general writing proficiency* of L2 learners' written products?
2. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *accuracy* of L2 learners' written products?
3. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *complexity* of L2 learners' written products?
4. What metacognitive strategies do L2 learners perceive in their actual writing performance?

1.5. Significance of the study

The present study hopes to make important contributions to the field of second language writing in a number of ways. The study has the potential to provide constructive insights on how strategies may influence L2 writers' performance on writing tasks. The focus on how metacognitive strategy and regulatory skills may be taught, whether in sequence or in combination, will contribute to pedagogical frameworks for designing classroom syllabuses. Thus, as this study involves the design of lesson plans and the creation of questionnaire instruments for data collection and analysis, L2 writing teachers and researchers will have access to the material and the taxonomy of these strategies for L2 writing instruction. The study therefore intends to explore directions for further research, syllabus design, and curriculum development in the field of academic writing.

A novel element in this study also concerns the recognition of written academic literacy in EFL contexts. By teaching metacognitive strategies to a number of EFL learners in Iran, this study makes one of the first attempts to conceptualise and empirically test the metacognitive strategy instruction construct as a situational variable. Hence, the study argues that due to the globalisation phenomenon, entering and taking part in disciplinary realities, particularly full participation in scientific communities, requires proficiency in both spoken and written literacies that goes beyond the mastery of reading technical texts and translation.

1.6. Outline of the thesis

The thesis consists of seven chapters. As an introduction, Chapter 1 included the background of the study, the statement of the problems, the aims of the study, and a summary of the contents in this thesis. Chapter 2 reviews the related literature on studies in L2 writing instruction on the one hand and provides a survey of studies that have investigated metacognitive strategies and pre-task planning in writing on the other hand. Chapter 3 outlines the research methodology of the two studies conducted in this project. The research context, the participants, and the instruments used within each study are first described. The chapter then proceeds with a description of the data collection and data analysis procedure. To answer the first three research questions, Chapter 4 reports the quantitative findings obtained from the writing texts in the first study and then provides a discussion of the main findings, plus further suggestions for the second study. In the same way, Chapter 5 presents the quantitative textual analysis for the second study, followed by an in-depth discussion of the findings. Chapter 6 presents the report of the think-aloud data analysis and discusses the results of the questionnaires administered to participants. Chapter 7 summarises the study by using the findings obtained by the mixed-method analysis and then considers some pedagogical and theoretical implications for L2 writing development. It concludes with a consideration of the limitations of the study and suggestions for future research.

CHAPTER 2. REVIEW OF THE LITERATURE

Chapter 2 presents a detailed review of the previous research findings and the implications associated with the three different but potentially commensurable research areas this study draws on: strategies-based instruction research, second language (L2) writing research, and pre-task planning research in SLA. In these three sections, the main findings of the studies to date are discussed and the issues which require further investigation are identified. Finally, connections among these three salient threads relevant to the current study are established.

2.1. Language learning strategy instruction

Strategies have been defined as "actions, behaviours, steps, or techniques students use, often unconsciously, to improve their progress in apprehending, internalising, and using the L2" (Oxford, 1994, p. 1). All learning strategies share a number of features:

1. They are controllable (Pressley & McCormick, 1995);
2. They enhance self-efficacy, the individual's perception that he can successfully complete a task or series of tasks (Bandura, 1997);
3. With some exceptions, they are not inherently good or bad, but have the potential to be used effectively (A. D. Cohen, 1998);
4. They are teachable and the teaching is more successful when they are taught explicitly and are tied to the tasks students are normally expected to accomplish (Oxford, 1990).

Griffiths (2008) synthesised a definition from among the definitions of language learning strategies in the review of literature. She defined strategies as activities which are consciously chosen by learners for the purpose of regulating their own learning. It is worth noting that other researchers (e.g. Dörnyei & Skehan, 2003) had already abandoned the term 'strategy' in favour of 'self-regulation'. This illustrates the long-standing controversy over the definition and the consideration of language learning strategies as conscious activities (A. D. Cohen, 1998; Griffiths, 2008) or unconscious thoughts and behaviours (Oxford, 1994).

To this date, various typologies of strategies have been offered. One of the most comprehensible classifications has been proposed by Oxford (1990) who categorized strategies into six major types: metacognitive, cognitive, memory, compensation, social, and affective. Each strategy type fosters a particular dimension of learning language as communication and can be explained and justified on the basis of theoretical principles of first and second language development. This part of the literature review introduces the issues considered important in language learning strategy instruction research and concludes with an account of issues in metacognition, and metacognitive strategy instruction in writing.

2.1.1. Strategy instruction research

The literature on learning strategies in second language acquisition emerged in the 1970s when researchers who were influenced by an increase in the cognitive linguistic theories accounted for the characteristics of good language learners (Naiman, Frohlich, Stern, & Todesco, 1978; Rubin, 1975; Stern, 1975). These seminal studies led to numerous other language learning strategy studies in proposing strategy instruction models and in validating their effectiveness. Strategies-based instruction (SBI) refers to a form of learner-centred language teaching that incorporates, explicitly or implicitly, style and strategy instruction activities into everyday classroom language instruction (A. D. Cohen & Dörnyei, 2002; Oxford, 2001). The role of SBI in promoting learner autonomy is widely recognised (A. D. Cohen, 1998; Nguyen & Gu, 2013). SBI serves autonomy or Self-Directed Language Learning (SDLL) through identification of various learning strategies in language learning and devising methods and innovations which help students to become more aware of strategies available to them, understand how to organise and use strategies effectively according to their learning styles, and learn when and how to transfer the strategies to new language learning. A large body of research supports the positive effects of SBI in language learning performance (Carrell, Pharis, & Liberto, 1989; Oxford, 1990). Oxford et al. (1990, p. 210) asserted that "strategy training can enhance both the process of language learning (the strategies or behaviours learners use and the affective elements involved) and the product of language learning (changes in students' language performance)." Cottrell (2003) proposed that through practice and instruction, learners' use of strategies can be automatised. (Dörnyei, 2005, p. 178) suggested that strategy training frameworks aim to achieve the following goals:

to raise learners' awareness about learning strategies; to encourage strategy use; to offer a number of relevant strategies for learners to choose from; to offer controlled practice in the use of strategies; to provide an analysis for students to reflect on their strategy use.

2.1.1.1. Strategy instruction models

There are a number of models developed for teaching language learning strategies, among which three instructional models are widely referred to in the literature: Styles and Strategies-Based Instruction (SSBI) model (A. D. Cohen, 1998), Cognitive Academic Language Learning Approach (CALLA) model (Chamot & O'Malley, 1986), and Grenfell and Harris's model (1999). All three aim to develop students' metacognitive understanding of the value of learning strategies and encourage them to adopt strategies which will improve their language learning and language proficiency.

Cohen's SSBI model has a series of components: 1) Strategy preparation, 2) Strategy awareness-raising, 3) Strategy instruction, 4) Strategy practice, and 5) Personalisation of strategies. In this model, the teacher acts as a diagnostician, language learner, learner trainer, coordinator and coach to make sure strategies are integrated into instructional materials. The sequence of instruction in the CALLA model of Chamot and O'Malley's, which was developed to increase the achievement of English-language learning students and other students who are learning through the medium of second language, is a five-phase recursive cycle for introducing, teaching, practicing, evaluating, and applying learning strategies. In the six-step model developed by Grenfell and Harris, the teacher raises learners' awareness, models and discusses the value of strategies, gives learners practice, sets goals, chooses appropriate strategies to attain goals, and, finally, the teacher and learners together evaluate the success of the action plan, set new goals, and a new cycle begins again.

To summarise Chamot's (2004, p. 21) key points, these instructional models of language learning strategy instruction share a number of features:

1) They all agree on the importance of developing students' metacognitive understanding of the value of learning strategies and suggest that this is facilitated through teacher demonstration and modelling.

2) All begin by identifying students' current learning strategies through activities such as completing questionnaires, engaging in discussions about familiar tasks, and reflecting on strategies used immediately after performing a task.
3) All suggest that the teacher should model the new strategy, thus making the instruction explicit.
4) All emphasise the importance of providing multiple practice opportunities with the strategies so that students can use them autonomously.
5) All suggest that students should evaluate how well a strategy has worked, choose strategies for a task, and actively transfer strategies to new tasks.

2.1.1.2. Explicit vs. Implicit strategy instruction

Two instructional methods for strategy use have generally been acknowledged: explicit and implicit strategy instruction. A few researchers are sceptical about the effectiveness of explicit strategy instruction and alternatively suggest that implicit strategy instruction should be given closer attention. Rees-Miller's (1993) rationale in calling teachers to exercise caution in the implementation of learner-strategy instruction in the classroom is the mixed results which have been obtained from various research studies. Before adopting any strategy instruction model, she maintains, teachers should take into consideration various complicating factors, including students' cultural differences, age, educational background, life experience, affective factors, varying cognitive styles, and students' and teachers' beliefs about language learning. Similarly, Eslinger (2000) asserts that leaners may have a natural tendency to use strategies without needing any explicit instruction of strategies.

There is now widespread agreement, however, on the importance and effectiveness of explicitness in strategy instruction curriculum in second language contexts (N. J. Anderson, 2005; H. D. Brown, 2002; A. D. Cohen, 1998; Nunan, 1996; O'Malley & Chamot, 1990; Oxford & Leaver, 1996; Scruggs, Mastropieri, Monson, & Jorgensen, 1985; Shen, 2003). In Chamot and O'Malley's (1986) CALLA model, instruction starts in highly explicit sequences and gradually fades in a way that students can begin to take on greater responsibility for selecting and applying appropriate strategies to new learning tasks. In a project conducted by Nunan (1996), integrating explicit instruction of language learning strategies into the classroom curriculum was achieved by incorporating a

conscious focus on strategies and encouraging learners to reflect on their learning processes and to develop skills in self-checking, monitoring, and evaluation. He concluded that "language classrooms should have a dual focus, not only teaching language content but also on developing learning processes as well" (p. 41). In his book Strategies for Success: A practical guide to learning English, H. D. Brown (2002) provided a practical resource guide for both students and teachers. The material includes a variety of ready-to-use pair- and group-work exercises for the four language skills, listening, speaking, reading and writing.

2.1.1.3. Integrated vs. Separate strategy instruction

Compared with the issue of explicitness/implicitness in strategy instruction, there is less agreement on whether strategy instruction should be taught separately or integrated into the regular language courses. Although there has been no comparative research on these two areas in second language acquisition, the literature largely supports the idea that integrated instruction provides students with opportunities to practice learning strategies with more authentic language learning tasks (Bachman & Palmer, 1996; Chamot & O'Malley, 1994; A. D. Cohen, 1998; Grenfell & Harris, 1999; Oxford & Leaver, 1996). In examining the effectiveness of a model of writing instruction for students with learning disabilities, MacArthur, Graham, Schwartz, and Schafer (1995) integrated word processing and strategy instruction to a process approach to writing instruction. After implementing the curricular model for a full school year, they observed that students in the experimental classes made greater gains in the quality of their narrative and informative writing.

On the other hand, those who opt for separate learning strategy courses argue that strategies learned within a language class are less likely to transfer to other tasks (Gu, 1996), and that it is practically easier to plan for one separate strategy course than to prepare all teachers to teach strategies (Vance, 1999; Weinstein & Mayer, 1986).

2.1.1.4. Strategy instruction in writing

Strategy instruction has been identified as an effective form of instruction for teaching writing (Englert, Mariage, & Dunsmore, 2006). The studies investigating the effectiveness of strategy instruction in both L1 and L2 writing contexts are robust (Cresswell, 2000;

Graham, 2006; Sengupta, 2000). One well-known and widely-used model of strategy instruction in writing developed by Graham and Harris (1993) is the Self-Regulated Strategy Development (SRSD) model. Students in this model are explicitly taught writing strategies through six stages, i.e. develop background knowledge, discuss it, model it, memorise it, support it, and individual performance.

The benefits of the impact of SRSD model on students' writing performance have been largely documented in a number of studies. De La Paz and Graham (2002) used the SRSD model to teach strategies for planning, drafting, and revising expository essays to 50 secondary school students. In the post-test phase of the study, students in the experimental condition were shown to have produced longer essays containing more mature vocabulary. In another experimental study, De La Paz (2005) used the model to teach argumentative essays to 62 secondary school students. Like De La Paz and Graham's (2002) study, this one also yielded positive results for the strategy instruction. The results indicated that the students in the experimental group wrote more accurate and more persuasive essays. Additionally, Hawthorne (2008) explored the impact of the SRSD treatment on the engagement and writing performance of a group of year 10 students in New Zealand, most of whom were classified as reluctant writers. The results indicated that the treatment had a significantly positive impact on reluctant writers' reported levels of motivation. In addition, the writing performance of both reluctant and non-reluctant groups improved significantly as a result of the treatment. Overall, in Graham's (2006) meta-analysis to examine the overall impact of studies using the SRSD model in the teaching of genres, ranging from narratives, expository, compare-and-contrast, to argument, most studies showed a large effect size of above .8, which provides a strong evidence for the substantial impact of the model on students' writing performance.

2.1.2. Metacognitive strategy instruction research

In the 37 years since Flavell (1976) coined the term metacognition, research studies have investigated the impact of metacognition on L1 and L2 teaching and learning (N. J. Anderson, 2002a, 2002b, 2007, 2008; Coutinho, Wiemer-Hastings, Skowronski, & Britt, 2005; Efklides, 2006; Efklides & Misailidi, 2010; Larkin, 2009; Prins, Veenman, & Elshout, 2006; Sánchez-Alonso & Vovides, 2007). Given the significant role of strategies in successful language learning, there are these questions of how important the

metacognitive components are in strategy instruction, how these components are related to the strategies themselves, and whether they can be teased apart from the strategy instruction. O'Malley and Chamot (1990) strengthened the importance of the role of metacognitive strategies when they stated that "students without metacognitive approaches are essentially learners without direction or opportunity to plan their learning, monitor their progress, or review their accomplishments, and future learning directions" (p. 8). According to Vandergrift (2002, p. 559), "Metacognitive strategies are crucial because they oversee, regulate, or direct the language learning task, and involve thinking about the learning process." Of the various categories of strategies identified through strategy research, N. J. Anderson (2005) hypothesised that the metacognitive strategies play a more significant role because once learners understand how to regulate their own learning through the use of strategies, language acquisition should proceed at a faster rate.

The importance of supporting learners through teaching metacognitive strategies has also been noted by Kolencik and Hillwig (2011). They claim that the whole purpose of teaching these strategies is "to increase students' self-awareness about what it takes them to learn" (p. 7). They believe in great variability in the growth, proficiency and refinement of these strategies. Thus they suggest that teachers or educators need to be aware of such developmental differences in metacognition so that they can "acknowledge, cultivate, and enhance the metacognitive capabilities of all learners" (p. 8). Likewise, according to Scruggs et al. (1985), increases in learning follow direct instruction in metacognitive strategies, but the independent use of these thinking strategies develops gradually.

The following section first presents the definitions, components and typologies of metacognition, then offers an examination of metacognitive language learning strategies assessment, and finally reviews the studies that have investigated metacognitive strategy instruction in writing.

2.1.2.1. Definitions of metacognition

Flavell (1976) first introduced the term metacognition to cognitive psychology to refer to one's knowledge concerning one's own cognitive processes or anything related to them, e.g. the learning-relevant properties of information or data (p. 232). He further explained that metacognition involves "active monitoring and consequent regulation and

orchestration of cognitive process to achieve cognitive goals" (p. 252). Later, Flavell and Wellman (1977) and Flavell (1979) included interpretation of ongoing experience, or simply making judgments about what one knows or does not know to accomplish a task, as other features of metacognition.

Within literature, metacognition has also been defined as simply "thinking about thinking" (N. J. Anderson, 2002a, p. 56), "thoughts about thoughts, knowledge about knowledge, or reflections about actions" (Weinert, 1987, p. 8), or an "awareness of one's own cognitive processes rather than the content of those processes together with the use of that self-awareness in controlling and improving cognitive processes" (Biggs & Moore, 1993, p. 527). Other researchers have referred to metacognition as evaluating cognitive states such as self-appraisal and self-management (A. L. Brown, 1987), and "knowledge about executive control systems" (C. Brown, Hedberg, & Harper, 1994, p. 3). According to Oxford and Crookall (1989, p. 404), metacognitive strategies refer to "the deliberate conscious and unconscious mental behaviours for directing and controlling cognitive strategy use for successful performance on a task." They describe metacognitive strategy use as the use of beyond-the-cognitive strategies that are used to provide executive control over the execution of the task. O'Malley and Chamot (1990) describe metacognitive strategy use as "thinking about or knowledge of the learning process, planning for learning, monitoring learning while it is taking place, or self-evaluation of learning after the task has been completed" (p. 231).

Considering these definitions, it should be clear that although the term metacognition has been part of the vocabulary of cognitive development and educational psychology for the last four decades, defining metacognition is anything but simple. One reason for this debated multiplicity of definitions, Livingston (1997) asserts, is that there are several terms interchangeably used in the literature to describe the same basic phenomenon (e.g. self-regulation, executive control, higher-order skills), or an aspect of this phenomenon (e.g. metamemory, comprehension monitoring). Yet, while there are some distinctions among these definitions, all unanimously emphasise the role of executive processes in the overseeing and regulation of cognitive processes.

2.1.2.2. Components of metacognition

Hardly anyone questions the reality or importance of metacognition. Yet, there is a considerable debate over the various components of metacognition represented in various studies. In examining the methodological and conceptual considerations of metacognition and learning, Veenman, Van Hout-Wolters, and Afflerbach (2006, p. 4) stated that:

> The domain of metacognition is one that lacks coherence… While there is consistent acknowledgement of the importance of metacognition, inconsistency marks the conceptualization of the construct... Apparently, more theoretical work needs to be done for attaining a unified definition of metacognition and its components.

Flavell et al. (2002) divided metacognition into metacognitive knowledge, and metacognitive monitoring and self-regulation. They then further divided metacognitive knowledge into three sub-categories: 1) Knowledge about persons; 2) Knowledge about tasks; and 3) Knowledge about strategies. The three facets of metacognitive knowledge have been identified as 'declarative', 'procedural', and 'conditional' knowledge (McCormick, 2003; Schraw, 1998). Declarative knowledge encompasses the three aforementioned sub-categories of metacognitive knowledge about the self, task, and strategies. It refers to one's knowledge about oneself as a learner and his/her strengths and weaknesses, and knowledge regarding the task and skills and strategies required to complete the task (Harris, Graham, Brindle, & Sandmel, 2009, p. 133). Procedural knowledge includes "information about how to successfully apply the various actions or strategies comprising declarative knowledge" (ibid, p. 228). Conditional knowledge "allows the writer to determine the appropriate conditions in which to apply the procedural and declarative knowledge" (ibid, p. 228). On the other hand, the metacognitive monitoring and regulatory skills of metacognition have been addressed by numerous researchers (e.g. Jacobs & Paris, 1987; Kluwe, 1987). This aspect of metacognition consists of the application of metacognitive strategies like planning (i.e. the selection of appropriate strategies and the allocation of resources), monitoring (i.e. analysis of the effectiveness of the strategies or plan being used), and evaluating (i.e. determining progress being made toward the goal which can result in further planning, further monitoring and further evaluation) (Harris et al., 2009). As Flavell et al. (2002) and N. J. Anderson (2002a) assert, all these components of metacognition work interactively.

Finally, there is a distinction between domain-general and task/domain-specific metacognition (Veenman & Beishuizen, 2004; L. J. Zhang, 2001). General metacognition transcends particular subject or content areas, such as setting goals, and can be instructed concurrently in different learning situations. Domain or task-specific metacognition refers to the metacognition which is applied in particular subjects or content areas and has to be taught for each task or domain separately, such as reading and writing. Although Schraw (2001) argues that metacognition is domain-general in nature, he acknowledges the substantial debate that exists over whether metacognitive knowledge is initially domain-general or domain-specific. In the former case, he asserts, as the general metacognitive expertise improves, one expects that these skills will span multiple domains. In the latter, as learners acquire more metacognitive knowledge in a number of particular domains, they are likely to flexibly construct general metacognitive knowledge and regulatory skills that cut across a wide variety of domains (p. 7).

2.1.2.3. Typologies of metacognitive strategies

A number of metacognitive strategy typologies have been developed over the past 30 years. In what follows, three of the most influential and widely-cited classifications by O'Malley and Chamot (1990), Oxford (1990), and N. J. Anderson (2007) will be presented. O'Malley and Chamot's (1990) classification of metacognitive strategies involves seven strategies and is differentiated to illustrate the planning, monitoring, and evaluating stages:

A. Planning:

1. Advance organisers: Previewing the main ideas and concepts of the material to be learned,
2. Directed attention: Deciding in advance to attend in general to a learning task and to ignore irrelevant distractors,
3. Functional planning: Planning for and rehearsing linguistic components necessary to carry out an upcoming language task,
4. Selective attention: Deciding in advance to attend to specific aspects of input or situational details that assist in performance of a task,
5. Self-management: Understanding the conditions that help one learn and arranging for the presence of those conditions.

B. Monitoring:

6. Self-monitoring: Checking one's comprehension or performance, or checking the accuracy and/or appropriateness of one's oral or written production while it is taking place.

C. Evaluation:

7. Self-evaluation: Checking the outcomes of one's language learning against a standard after it has been completed.

The above seven subcategories of metacognitive strategies can be further classified according to the three types of cognitive processes: advance organisation, online organisation, and post organisation. They do not move in a linear fashion from the first to the seventh strategy. More than one metacognitive strategy may be occurring at a time during a learning task and some strategies like selective attention may function at the advance or online organisations stage.

In her landmark book Language Learning Strategies: What Every Teacher Should Know, Oxford (1990) drew a distinction between direct and indirect strategies, and considered metacognitive strategies among indirect strategies. As she put it (p. 17), metacognitive strategies help learners to regulate their learning through planning, arranging, focusing, and evaluating their own learning process. Oxford's taxonomy of metacognitive strategies consists of the following three stages (p. 137):

A. Centring your learning:

1. Overviewing and linking with already known material
2. Paying attention
3. Delaying speech production to focus on listening

B. Arranging and planning your learning:

4. Finding out about language learning.
5. Organising
6. Setting goals and objectives
7. Identifying the purpose of a language task
8. Planning for a language task
9. Seeking practice opportunities

C. Evaluating your learning.

10. Self-monitoring

11. Self-evaluating

Centring one's learning in Oxford's typology involves online organisations strategies, and arranging and planning one's learning is an advance organisation stage. The evaluation of one's learning can be regarded as relevant to both online and post organisation stages since self-monitoring and self-evaluating may be engaged both in the course of task execution and after the task completion. As reported, the two aforementioned classifications of Oxford's, and O'Malley and Chamot's, albeit expressed in different terms, share some levels of common ground in subdividing metacognitive strategies in terms of the three types of cognitive process.

As can be seen in Figure 2.1, the third taxonomy belongs to Anderson's model (2007) which proposed five main components for metacognition. These include: 1) preparing and planning for learning, 2) selecting and using learning strategies, 3) monitoring strategy use, 4) orchestrating various strategies, and 5) evaluating strategy use and learning. He explained that these components are not used in a linear way, but rather they interact with one another. Thus he highlighted how providing opportunities for learners to reflect upon their learning strategies and orchestrate the use of more than one strategy available to them facilitates the improvement of strategy use and empowers second language learning.

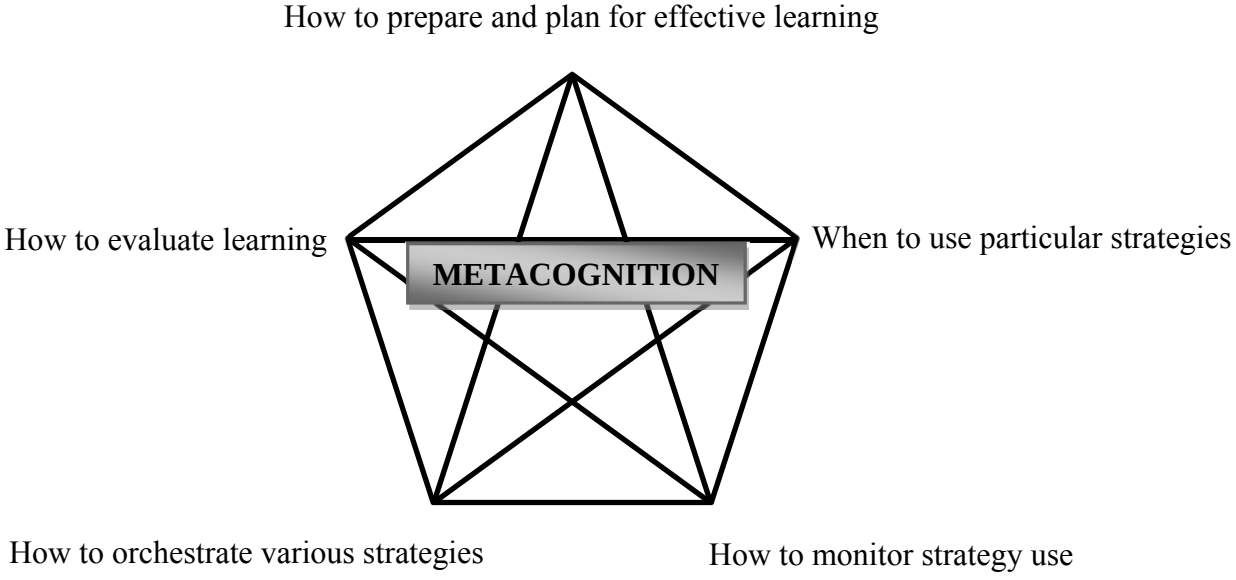


Figure 2.1 Anderson's model of metacognition (2007)

This classification of metacognitive language learning strategies can similarly be reflected within the three general aspects of metacognition. Engaging in preparation and planning in relation to a learning goal can be directly grouped into advance organisation which occurs before a learning task starts. The next three strategies, i.e. selecting and using learning strategies, monitoring strategy use, and orchestrating various strategies, represent the strategies involved in online organisation. Finally, the evaluation of strategy use and learning may be either grouped into online organisation stage within the course of task execution, or into post organisation stage after the task completion. Table 2.1 provides a summary of these three typologies of metacognitive strategies and how they can be reflected within advance, online and post organisation stages.

Table 2.1 Summary of three typologies of metacognitive strategies

	O'Malley & Chamot (1990)	Oxford (1990)	Anderson (2007)
Advance Organisation	Advance organisers Directed attention Functional planning Selective attention Self-management	Arranging and planning one's learning	Preparing and planning for effective learning Evaluating strategy use and learning
Online Organisation	Directed attention Functional planning Selective attention Self-management Self-monitoring	Centring one's learning	Deciding when to use particular strategies Knowing how to monitor strategy use Learning how orchestrate various strategies
Post Organisation	Self-evaluation	Evaluating one's learning	Evaluating strategy use and learning

2.1.2.4. Assessment of metacognition

Ongoing research and application of metacognition in language learning require various measuring instruments. Many methods of assessment have been used to describe and measure metacognition, both in educational research and practice. These include offline methods (i.e. whether the measurement takes place before or after the task performance) like questionnaires (Pintrich & De Groot, 1990; Schraw & Dennison, 1994; Thomas, 2003; Thomas, Anderson, & Nashon, 2008), oral interviews (Zimmerman & Martinez-

Pons, 1990), stimulated recall (B. Van Hout-Wolters & Duffy, 2000), journals (N. J. Anderson, 2007; Goh, 1997), or online methods (i.e. whether the assessment takes place during the learners' learning) like thinking aloud protocols (Afflerbach, 2000; Veenman, Elshout, & Groen, 1993), observations (Veenman & Spaans, 2005), online computer logfile registration (Veenman, Wilhelm, & Beishuizen, 2004), and eye-movement registration (Hacker et al., 2009; Kinnunen & Vauras, 1995). Each of these assessment methods have their pros and cons and can be beneficial depending on the metacognitive knowledge or skill component that the researcher intends to assess. According to Schellings (2011), the distinction in online versus offline measuring methods indirectly relates to task-specific or general learning discussed in section 2.1.2.2 Components of metacognition. In a similar vein, Schellings argued that whereas online measures can be targeted at the assessment of task-specific metacognitive strategies during task performance, offline measures are more bound to the measurement of general metacognition.

Because of the employment of questionnaires and think-aloud protocols in the current study, a brief review of the studies examining metacognitive strategies via these research instruments is provided in this section. According to Schellings, van Hout-Wolters, Veenman, and Meijer (2012), the existing metacognitive questionnaires in educational research differ in their level of specificity, goals, content, aimed population, reliability and validity. Many questionnaires are aimed at assessing general strategies irrespective of the learning task or content (e.g. Approaches to Studying Inventory (ASI), Entwistle and Ramsden (1983); Study Process Questionnaire (SPQ), Biggs (1987), Learning and Study Strategies Inventory (LASSI), Weinstein, Schuke, and Palmer (1987); Motivated Strategies for Learning Questionnaire (MSLQ), Pintrich, Smith, García, and McKeachie (1991). A number of specific questionnaires have also been constructed which aim at measuring particular aspects of metacognitive strategies in educational research and practice (e.g. Metacognitive Awareness Inventory (MAI), Schraw and Dennison (1994); State Metacognitive Inventory (SMI), O'Neil Jr and Abedi (1996); Self-efficacy and Metacognition Learning Inventory-Science (SEMLI-S), (Thomas et al., 2008).

A large number of empirical studies have investigated the effects of different verbalisation methods on both learning processes and learning outcomes. In examining the differences in the self-regulation processes between successful and less successful students, F. De

Jong (1987) used two different think-aloud methods to assess metacognitive skills. In one, the participants were instructed to think aloud constantly, while in the other, they were prompted to verbalise their thoughts only at certain marked points. De Jong concluded that while both these online methods could predict learning outcomes more accurately than the scales of the offline questionnaire, the instruction of constant thinking aloud is more recommended for assessing metacognitive skills. However, in neither Veenman's (1993), nor Crain-Thoreson, Lippman, and McClendon-Magnuson's (1997) studies did thinking aloud procure different learning outcomes in groups which thought aloud constantly compared to the groups which thought aloud markedly, or learned silently.

Schellings (2011) contrasts the strength of think-alouds over questionnaires and offers some explanations for the low correlations between the two measures (pp. 95-96). First, learners may have forgotten some learning activities and consider some to be too unimportant to mention. Secondly, some learning activities may be mentioned that did not actually take place. Thirdly, a learner may be unaware of some of the learning activities that he or she executed or may not be able to reflect in a way necessary for the correct completion of the questionnaire. Finally, social desirability impacts the completion of the questionnaire. He then offers two alternative explanations, the first being the learning strategies measured and the second being the learning task to which the measurement is related. Firstly, he explains that questionnaires may, for instance, be aimed at measuring learning strategies which are different in content from the strategies that may be analysed via think-aloud protocols. Secondly, he claimed that many of the questionnaires in research and practice are general, whereas think-aloud methods measure very specific learning tasks.

Learning strategy questionnaires have been compared with other online or offline methods and measures. Inter-correlations ranging from -0.07 to 0.60 have been reported when relating general questionnaires to other general questionnaires (Muis, Winne, & Jamieson-Noel, 2007; Sperling, Howard, Staley, & DuBois, 2004). When different subscales of general questionnaires have been correlated with the subscales of task-specific questionnaires, moderate correlations have been reported (e.g. a range of 0.24 to 0.63 in Samuelstuen and Bråten (2007); a correlation of 0.49 in Veenman (2005). When comparing online methods with each other in multi-method research measuring metacognitive strategies, both Veenman (2005) and Bernadette Van Hout-Wolters (2009)

reported high correlating results. For instance, Veenman (2005) reported correlation range of 0.64 to 0.89 when think-aloud method was compared with observation of behaviour and the logfile method. Because of such high correlations, Schellings (2011) encouraged "the possibilities of using different online methods within the same situations and for the same goal" (p. 95).

Low or moderate correlations are regularly reported between think-aloud measures and questionnaires (e.g. Veenman, 2005). In her study measuring different kinds of assessment methods of learning strategies and their usefulness in education and research, Bernadette Van Hout-Wolters (2009) reported variable correlations ranging from -0.07 to 0.42. Yet, Van Hout-Wolters differentiated the correlations between think-aloud measures and general questionnaires (reaching up to 0.22) and think-aloud and task-specific questionnaires (reaching up to 0.42). In a case study conducted by Schellings (2011), however, he found a higher correlation between the two measures ($r = 0.51$) and claimed that the reasons for lower correlations in the previous research may be that (1) different learning styles may be measured in such multi-method studies and (2) the measuring methods may be aimed at different learning tasks. In this way, he set two methodological prerequisites for comparing the instruments: the questionnaire and the think-aloud method should be aimed at the same learning task and at the same learning activities.

2.1.2.5. Metacognitive strategy instruction in writing

Strategy instruction is considered metacognitive only when it actively engages students in understanding their learning (Flower, 1994). An important instructional implication of the focus on metacognition is how metacognitive skills are learned within the context of realistic problem-solving situations. After some fruitful research on the impact of metacognition in reading and reading instruction, the research received more attention in the 1990s to explore the application of metacognition in writing and writing instruction. The foregoing section explores the research on metacognitive strategy instruction in writing. Among the first attempts to point out the significance of metacognition in L2 writing was Devine's (1993) who suggested there was a need to "provide explicit information about the role of metacognition in second language writing" (p. 118). Similarly, Mayer (1998) asserted that an instruction which focuses only on a basic skill appears to be incomplete, but when combined with metacognitive instruction, it can help

students to benefit from the instruction which is sensitive to the metacognitive demands of the task. According to Hayes and Flower's (1986) analysis of the writing process, composing an essay entails planning, translating and reviewing. Carter (1990, p. 281) noted that the learning of writing requires two important inputs: the acquisition of general knowledge about writing, and domain-specific local knowledge. Akin to this view of writing, metacognition, a concept including "both metacognitive knowledge and the processes of regulation and control of cognition" (Larkin, 2009, p. 150) is thus fundamental to skilful writing.

Englert et al. (1991) evaluated the effectiveness of a program called Cognitive Strategy Instruction in Writing, which emphasised cognitive instruction and metacognitive learning through combinations of strategies in elementary education. The results indicated that the cognitively and metacognitively instructed writers were more sensitive to the audience needs and could describe their planning, writing and revision processes more explicitly. Another extended example of metacognitive instruction in L1 writing is Smagorinsky (1991) who compared the effectiveness of three instructional treatments among high school students: models (declarative knowledge of form), general procedures (declarative knowledge of form plus general procedural knowledge related to content plus procedural knowledge related to form), and task-specific procedures (declarative knowledge of form plus task-specific procedural knowledge related to content plus procedural knowledge related to form). The results of think-aloud protocol analysis showed that students in the task-specific procedures scored the highest on purposeful composing and critical thinking. General procedure treatment was more effective than the models treatment only in linking ideas according to particular task constraints. In sum, De La Paz (2007) found that cognitive strategy instruction programs were effective in increasing writing performance. Such programs, she asserted (p. 250), employ the following five procedures in their explicit instructions:

> Teachers routinely provide think-aloud demonstrations; instruction is initially teacher-directed, with gradual fading of instructional scaffolds (i.e., prompts or supports such as dialogue); students work collaboratively and then independently to master the target strategies; responsibility for regulating the use of strategies is transferred from the teacher to the students; and the use of mnemonics to help students to remember the sequence of steps for executing each strategy – for each letter in the mnemonic cues, students are to engage in a cognitive or metacognitive

act (such as self-regulation) as part of the overall sequence of elements, resulting in the overall execution of a strategy.

Metacognition can be viewed as mental models of cognition, and these may vary from learner to learner (Devine et al., 1993). Devine et al.'s study attempted to provide information about the role of metacognitive knowledge in L2 writing by comparing L1 basic writers' and L2 writers' cognitive models of the writing process and their writing performance. They found that English L2 writers experienced tension between writing in a grammatically correct manner while also retaining their personal voice; thus they had cognitive models which contained conflicting demands in writing. Interestingly, Devine et al. found no such conflict between the demands of grammatical correctness and other concerns in L1 writers' writing. Furthermore, results of their research suggested that there were potential links between the cognitive models of writing employed and the actual performance on writing tasks. These authors concluded that a key goal of teachers of second language writing should be to increase their students' metacognitive knowledge. In doing so, teachers may need to determine whether students hold appropriate notions about the nature of writing. Alternatively, they may also design activities to address the gaps in students' knowledge base of person, task or strategy variables.

In assessing students' attitudes towards writing, perceptions of competence as writers, and declarative, procedural, and conditional knowledge about composing, Graham, Schwartz, and MacArthur (1993) found that normally achieving students and students with learning disabilities (LD) hold different conceptualisations to planning, revising and editing. As regards planning, students without LD placed greater emphasis on writing processes (e.g. organising information) which are required for task completion, whereas students with LD stressed information generation more than other activities. When queried about how to revise and edit a text, students without LD were more likely to emphasise substantive considerations, such as rewriting, rearranging, audience awareness, while students with LD stressed the surface-level mechanical features, such as neatness, spelling, punctuation, capitalisations, and grammar errors. Similarly, S. C. Lin, Monroe, and Troia (2007) asserted that a careful attention should be directed to the students' developmental level during writing instruction. To extend the understanding of writing knowledge development in school-age children, they examined students' developmental patterns for metacognition about writing. The results indicated that differences in various types of

writing knowledge were evident across grades and ability levels. They showed a progressive pattern by illustrating that novice and struggling writers focused on the physical product and local meaning, while expert writers placed more emphasis on global aspects, such as meaning and communication with an audience.

Macaro (2001) conducted a study of five-month writing strategy instruction with six classes of secondary students of French in England. The instruction involved a variety of writing strategies which included the metacognitive strategies of advance preparation, monitoring, and evaluating. The results of the post-test writing tasks of the experimental groups showed significant gains in the grammatical accuracy of their writing. Moreover, the analyses of the pre- and post-test questionnaires and think-aloud interviews indicated that the experimental groups changed in their approach to writing, became less reliant on their teacher, more selective in their use of dictionary, and more careful about their written work.

He's (2005) study of adult Taiwanese EFL writers uncovered some of the metacognitive processes underlying writing in second language: planning, evaluation of the organisation and ideas of written drafts, revision of spelling and grammar. This study showed that, when encountering writing difficulties, participants turned to previous knowledge, mental schemata, and outside references. In another study, Mayer (2001) indicated that although students may have the cognitive skills needed for writing such as translating, i.e. transforming ideas into written text, they may be unable to use these skills productively. He supported a direct strategy instruction to teach students how to systematically generate a writing plan and how to review and revise what they have written in light of their plan. In a study applying metacognition in EFL writing instruction in China, Xiao (2007) proposed guidelines for EFL instructors to make metacognitive teaching an integral part of writing instruction. Xiao favoured the idea that EFL writing instructors should teach with metacognition (reflecting upon their own teaching and equipped with both metacognitive strategic knowledge and executive management strategies) and for metacognition (helping learners to build up metacognitive knowledge and experiences) (p. 19).

A number of existing studies within the last two decades have focused on those writing strategies which support the drafting and revision stages of the writing process (Cresswell, 2000; Cumming, 1995; Cumming & So, 1996; Ransdell, Lavelle, & Levy, 2002;

Sengupta, 2000; Zhu, 1995). Among these, Cresswell's and Sengupta's studies will be reported in detail here. In order to help develop responsible self-monitoring writers, Cresswell (2000) implemented a three-stage programme of procedures, which involved 1) raising awareness of the composition process and product, 2) writing marginal annotations, and 3) evaluating annotations. Over a period of four weeks, seven adult Italians were trained to write four self-monitored compositions, the chosen genres being opinion articles, and letters to newspaper editors. The first two compositions served for instruction purposes. The written annotations on the margins of the last two compositions were analysed in terms of the attention paid to global content and organisation. Ultimately, Cresswell proved the instruction programme to be effective in increasing the students' attention to both linguistic concerns such as grammar and spelling, and to reviewing the content and organisation necessary for successful writing. He concluded that the self-monitoring technique with learning instruction is capable of increasing students' autonomy and teacher responsiveness to individual students' needs as they learn how to write (p. 243). In an exploratory comparative-design study, Sengupta (2000) investigated the impact of explicit revision-strategy instruction on learners' performance and perceptions about writing. The holistic analysis of the composition scores of the experimental groups, who were trained to revise their drafts in a more reader-friendly way, proved to be greater than their peers in the control group. The data obtained from questionnaires and interviews in this study also showed students' positive perceptions towards being instructed to pay attention to the writing purpose and their intended readers in revising their texts.

It is worth noting that in the later models of metacognition, social factors such as collaboration or cooperation are acknowledged as influences on metacognition involved in writing (Galbraith & Rijlaarsdam, 1999; Gao & Zhang, 2011; Larkin, 2009; Wegerif, Mercer, & Dawes, 1999; Zimmerman & Risemberg, 1997). Based on a socio-cultural constructivist view of learning, Larkin (2009) worked collaboratively with teachers of six different year 1 and year 2 classes across five primary schools in UK to encourage facilitation of metacognition throughout all activities around the three areas of Process Talk, Presentation Talk, and Reflective Talk. Later, during the main phase of the project, she observed and reflected upon the lessons. This study is different from other studies of metacognition in the classroom (e.g. A. L. Brown, 1987; Kluwe, 1987) in the approach it has taken. Unlike other studies which were based on the individual development of

metacognition either alone or within group settings, Larkin's study focused on the social factors in the collaborative group work influencing metacognition. The study explored whether metacognition may be socially constructed or mediated between pairs, and may consequently impact the product of the partnership. Although she later acknowledged the difficulty of analysing group interactions through the theories of metacognition, she asserted this complexity of the two phenomena of metacognition and social relationships may in fact prove fruitful in developing new theories of metacognition and in providing guidance for pedagogy.

Summary

Instruction in metacognitive awareness and control of writing has focused on the impact of cognitive processes, specifically planning and revising, on text structure knowledge. Sitko (1998) argued that students need regular metacognitive instruction in how to evaluate and integrate strategies into their own repertoire so that they can control the complex cognitive and social processes involved in producing a text. According to Graham and Perin's (2007) meta-analysis of writing instruction for adolescent students, explicit strategy instruction in planning, revising, and editing text combined with meaningful practice opportunities enhances adolescents', particularly struggling students', knowledge and self-regulation and has a moderate to strong impact on their writing quality. Yet to be determined is if and how this kind of metacognitive instruction can enhance the quality of L2 learners' academic writing performance.

2.2. Second language writing

Writing is a complex and multifaceted activity which involves the writers to not only consider their grammatical ability and lexical knowledge and control, but also maintain a focus on their sense of what a unified subject is, their factual knowledge about the subject, their understanding of the world and their place in it, their audience perspectives and needs, etc. (Bereiter & Scardamalia, 1987; Harris & Graham, 1992). In fact, every writing task, as Grabe and Kaplan (1996) have asserted, is situated in a rhetorical context, involving complex interrelationships among various elements of writing: the writer, the reader, the text and the context. Furthermore, writing in L2 involves additional challenges for learners at different levels of proficiency; writing for different purposes, and in different writing contexts. Therefore in reviewing the literature on the development of L2

writing research, the discussion will hinge on the distinctions between the two clusters of L1 and L2 writing research although the research orientation towards L2 writing still largely follows similar strands to those used in L1 writing research (Grabe & Kaplan, 1996; Silva, 1993). In particular, the discussion on writing research in this section will include the approaches to writing research, L1 and L2 models of writing processes, the academic literacy for L1 and L2 writers, and the assessment of writing ability.

2.2.1. Approaches to writing research

According to Hyland (2010), there are three different types of approaches to researching writing, which are actually three different epistemological views of what constitutes writing. In this section, for each of these approaches related to the text, writer, and reader, a brief summary of what these markedly different views involve, and their development and research trends will be provided. Then a number of selected studies which have over the last decades impacted the theory development and the instructional practices in classrooms will be reviewed.

The first approach is text-oriented research approach which views writing as "a product, an artefact of activity which can be studied independently of users by counting features and inferring rules" (ibid, p. 191). A number of measures used extensively in studies following this perspective include syntactic complexity (ratio of clauses to some general unit, e.g. T-units, c-units, AS-units) or grammatical accuracy. This approach was developed in the 1950s and was heavily influenced by the school of structural linguistics. Accordingly, the production of linguistic aspects of writing, and how they are combined to create a good piece of text are considered significant within this approach (Grabe & Kaplan, 1996). However, Hyland (2010) doubted these measures could be the principal features of writing development or that they may be the best indicators of good writing. Hyland (2010), and Grabe and Kaplan (1996) pointed out that the perspective of viewing texts in this way ignores their role as a source of communication and how they function as a writer's rhetorical response to the needs of an audience in a particular communicative setting.

The second writing research approach is writer-oriented in which, as the name suggests, the main foci are the actions of writers and their cognitive performance, rather than the

surface structures and features of text as in the first approach. To elaborate the writing processes and activities that writers engage in to create and express meaning, this approach draws on models and tools of cognitive psychology and the analogy with artificial intelligence and computer processing. The methods employed in this approach are varied in the literature and include think-aloud protocols (e.g. Smagorinsky, 1991), retrospective interviews (e.g. Nelson & Carson, 1998), and task observation (e.g. Bosher, 1998). Psycholinguistic writing research has typically used data collected from think-aloud protocols to identify the strategies used by writers and to basically model the mental systems responsible for the production of the written texts. According to Hyland (2010), this perspective to writing research is significant in the sense that it gives greater emphasis to individual experiences of writers and how they understand local features of the context they deal with as they write. In response to the inadequacies of the traditional textual and rhetorical models of writing, this perspective thus shifted writing research to a direction in which writers are allowed to make their own meanings of their world based on their interest (Hyland, 2002). In this viewpoint, however, there is again no consideration of context in which the text is produced and no attention to the communicative function of texts, particularly with regard to "evoking a social milieu which intrudes upon the writer and activates specific responses to recurring tasks" (Hyland, 2010, p. 194).

This leads us to the most current writing research which takes a reader-oriented approach to writing by exploring the ways writers construct and understand their intended audience. This perspective, on the one hand, has its root in Bakhtin's (1981, 1986) notions of heteroglossia and dialogism which consider an interaction between the writer, reader, and their sociocultural worlds, as an ongoing, dialogic activity and as he put it, "each utterance refutes, affirms, supplements, and relies on the others, presupposes them to be known and somehow takes them into account" (p. 91). Furthermore, the concept of audience explains 'Rogerian argument' which takes its name from Carl Rogers and his work in psychology, and particularly depends on the writer's understanding of and finding common grounds with an audience (Rogers, 1970). Central to this recent perspective in researching writing are the studies into genre writing which, according to Hyland (2010), "seeks to show how language forms work as resources for accomplishing goals, describing the stages which help writers to set out their thoughts in ways readers can easily follow and identifying salient features of texts which allow them to engage effectively with their readers" (p. 196). Classroom applications of genre, known as genre-based approaches to writing, are

generally situated within the macro-purposes of the whole text, rather than the semantic micro-functions of individual words and sentences (Hyland, 2002; B. Lin, 2006). The benefits of a genre-based pedagogy of writing are well-documented in the literature. For instance, Paltridge (2002), in detailing the implementation of the genre approach to teaching English for Academic Purposes (EAP), maintained that a genre-based curriculum helps students whose backgrounds are distant from the conventions and expectations of academic writing to participate more successfully in academic discourse. Johns (2002), in her edited book Genre in the Classroom: Multiple Perspectives, argued that the explicit teaching of genres would facilitate the learning of writing. Further, B. Lin (2006) pointed to the evidence that the application of a genre-based pedagogy in an EFL course for Business undergraduates in a Japanese university improved students' macro-structure of their academic writing, and enhanced their content organisation at the paragraph level.

The shift towards the interdisciplinary dimension of context-oriented research in the 1980s has been characterised by three prominent but not mutually exclusive perspectives: social interactionism, social constructionism, and sociocultural perspective. The social interactionist perspective of writing takes into consideration both the social interactions and the individual writers' internal processes to interact with readers (Martin Nystrand, 1989). The social constructionist framework embodies two strands of research – the cognitive aspects of writing process, and Vygotsky's sociocultural theory featuring the social dimensions of text production, including audience, other texts and negotiated understandings (Flower, 1994). This perspective highlights the role of writing teachers in offering guidance via direct modelling and explicit instruction until the cognitive process of writing gets internalised into the writer's mind. The sociocultural view of writing not only recognises the cognitive internalisation of writing processes, but also embodies the externalisation of writing which results from the interaction with the historical, cultural, and social identities individual writers bring to writing, the social world in which the writing occurs, the peer and teacher interactions surrounding the writing, and the classroom organisation, including the curriculum and teacher or school pedagogical decisions (Englert et al., 2006; Prior, 2006).

2.2.2. Models of L1 and L2 writing process

The similarities and differences between L1 and L2 composing processes have been largely pointed out in the literature (Kelly, 1986; Moragne e Silva, 1989). Leki, Cumming, and Silva (2008) argue that apart from similar composing strategies in L1 and L2, difficulty with language, more at the rhetorical and stylistic level than in sentence grammar, and efficiency in reaching goals differ between L1 and L2. The schematic backgrounds and writing experiences of L2 learners are among other differences. As pointed out by some scholars in writing research (e.g. Cumming, 2005; Grabe, 2001; Barbara Kroll, 2003), there are no universally accepted L2 theories of writing. However, there is general acceptance that the available L1 models in writing research (Bereiter & Scardamalia, 1987; Grabe & Kaplan, 1996; Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1996; Zimmerman & Risemberg, 1997; Zimmermann, 2000) are broadly applicable to L2 writing.

Hayes' (1996) model of writing, which is the revised version of the earlier Hayes and Flower's (1980) model, consists of three cognitive subcomponents, namely planning, text generating and revising. In this revised model, he also postulates two main elements of the task social and physical environment, and the individual motivational and affective constructs in writing. These models have been two of the most influential and frequently cited ones in L1 and L2 writing research and instruction.

Another model of writing which details the critical differences in the processes used by novice and expert writers is developed by Bereiter and Scardamalia (1987). They divide the writing process into four sub-processes: (a) a mental representation of the task, (b) problem analysis and goal setting, (c) problem translation, and (d) resultant knowledge telling. In this model, struggling writers are described as those involved in simple knowledge telling by basically retrieving content from long-term memory and writing it down. Expert writers, in contrast, tend to retrieve and transform content and discourse knowledge to plan text content in accordance with rhetorical, pragmatic and communicative dimensions. These processes operate in close interaction and are then elaborated in writing. However the resulting written text is continuously analysed and further content and rhetorical planning occurs. This complex process has strongly

influenced practices in strategies instruction, particularly when it is intended to develop metacognitive abilities (Harris et al., 2009).

Another model which is selected as the basis of the present study is developed by Kellogg (1996). In his model of writing production, Kellogg noted that it is reasonable to assume that processes involved in oral and written productions have much in common. Kellogg's model distinguishes three basic systems, formulation, execution and monitoring, with each system having two principal components. Formulation consists of (1) planning, where the writer establishes goals for the writing, thinks up ideas related to the goals and organises these to facilitate action, and (2) translating, where the writer selects the lexical units and syntactic frames needed to encode the ideas generated through planning and represents these linguistic units phonologically and graphologically in readiness for execution. Execution entails (3) programming, where the output from translation is converted into production schema for appropriate motor system involved (e.g. handwriting or typing) and (4) executing the actual production of sentences. And monitoring requires (5) reading, where the writer reads his or her text and (6) editing, which can occur both before and after execution of a sentence and can involve attending to micro aspects such as linguistic errors and/or macro aspects such as paragraph and text organisation.

In short, to this date there is no single model of second language writing that can adequately explain learning to write in a second language (including English as a second or foreign language). As Cumming and Riazi (2000) have pointed out, the information on "how people actually learn to write in second languages" is scarce (p. 57). Therefore, the information obtained from L1 writing models is considered a starting point for researchers and practitioners within the field of L2 writing. Depending on the specific context of pedagogical writing practices, the researchers may examine how an L1 model interacts with the L2 essential considerations of curriculum content, learners and instructional contexts.

2.2.3. Academic literacy for L2 writers

Prior to the 1970s, the teaching of L2 writing focused on the accurate reproduction of text models at sentence-level or discourse-level structures (Kaplan, 1966; Pincas, 1962). Drawing from cognitive linguistics, the theories of writing research in the 1970s were

expanded to cover the process-based approach on the one hand (Emig, 1971; Flower & Hayes, 1981; Zamel, 1976) and the development of English for Specific Purposes (ESP) as well as English for Academic Purposes (EAP) on the other hand. The notion of writing as process shifted the perspective of writing as the properties of texts and the reproduction of pre-learned syntactic and discourse structures to the process of writing and the development of organisation as well as meaning. Furthermore, the introduction of writing as language use in their specific contexts largely attempted to describe various aspects of academic writing literacy including features of academic genre, academic writing needs, and specific tasks required in courses across different disciplines (Johns, 2003). Leki (2007) defines academic literacy as "the activity of interpretation and production of academic and discipline-based texts" (p. 3). Academic literacy has also been well defined as "the capacity to undertake study and research, and to communicate findings and knowledge, in a manner appropriate to the particular disciplinary conventions and scholarly standards expected at university level" (University of Western Australia, 2005). According to N. Murray (2010), there is now widespread agreement that both domestic and international students with either English- or non-English-speaking backgrounds need to learn these academically valued writing skills. He asserts that although all these students enter university with some degree of literacy, they still need to learn the literacies they will require within the context of their discipline area, embedded within the curriculum of their undergraduate studies.

Matsuda (2003b) provides a historical perspective on the interdisciplinary relationship between composition studies and second language writing studies. In the latter part of the twentieth century, he reports, the field of basic writing emerged to accommodate both native-born and foreign-born students into writing courses. This placement of L2 writers in basic writing courses initiated a flurry of contentions, with many arguing that English native speaker students and second language learners have differing needs and thus should be taught separately and some arguing that these two groups of students could benefit being taught together. The common belief focused on the effects of L1 writing proficiency on L2 writing or the impact of L1 literacy instruction on L2 literacy development and not the other way round. However, further second language studies (e.g. Shaughnessy, 1976; Silva, Leki, & Carson, 1997) asserted that many theoretical and pedagogical insights from L2 studies, including error analysis, vocabulary lists and controlled composition, were applicable to basic writing instruction. This exchange of insights between the two types of

studies led researchers to recognise the multidisciplinary nature of L2 writing research and pedagogy (D. M. Johnson & Roen, 1989; Barbara Kroll, 1990; Matsuda, 2003a). Consequently, the field of second language writing evolved into an interdisciplinary field drawing insights from both composition studies and second language studies.

According to OECD, the number of international students has increased from 0.8 million in 1975 to 4.1 million in 2010. This increase in the number of internationally mobile students seeking to pursue their tertiary education in both English- and non-English speaking countries has led to a sub-field of L2 writing research referred to as L2 academic literacy. The growing interest in aspects of academic writing in English-medium educational institutions is attested to by the great deal of L2 writing research in this area and the specialised academic writing courses offered at such institutions to tertiary-level students. These students are expected to meet the academic literacy requirements in a language they did not acquire as their first language (Barbara Kroll, 2003). In addition to the linguistic demands L2 writers encounter, developing L2 literacy and academic competence is influenced by their cultural, educational, family and personal backgrounds (Leki, 2007). A large number of studies have examined literacy practices and cross-cultural responses to writing (e.g. Canagarajah (2002) in Sri Lanka; Ramanathan (2003) in India; Reichelt (2003) in Germany). The proliferation of this array of research has added insight into how cultural backgrounds shape L2 writing literacy practices and preferences. In short, the large variety of dimensions involved in the academic literacy experiences of L2 writers prompts researchers and practitioners equally to improve second language writing instruction at tertiary level.

2.2.4. Assessment of writing ability

As Hamp-Lyons (1995, p. 759) has stated, writing is a complex and multifaceted activity. When writing teachers or testing agencies assess student writing, they engage in even another complex and multifaceted activity, which is judging another person's text. There has been an ongoing discussion on how the assessment of writing can be restructured to better assess the students' level of writing competence and to better facilitate the teaching of writing (e.g. Hyland, 2003; Shaw & Weir, 2007). According to Hyland (2003, p. 214), there are five main reasons for evaluating learners which can also be applied to writing assessment: placement (i.e. to allocate students to appropriate classes), diagnostic (i.e. to

identify students' writing strengths and weaknesses and to propose remedial courses), achievement (i.e. to enable students to demonstrate the writing progress they have made in their course), performance (i.e. to give information about students' ability to perform particular writing tasks), and proficiency (i.e. to assess a student's general level of competence, and to provide certification for employment, university study, etc.). In what follows, a brief overview of different types of writing assessment and rating scales will be presented.

2.2.4.1. Types of writing assessment

Writing has historically been assessed in two different ways: indirect assessment and direct or performance assessment. The former involves the utilisation of either selected-response items (e.g. multiple-choice) or limited production items (e.g. short answers) to measure some of the components of writing skill, whereas the latter requires test-takers to produce an actual sample of writing (Weigle, 2002).

Each of these two types of assessments offers some advantages and disadvantages over the other type. Indirect assessment has been frequently used in large-scale assessments, because of the inclusion of more items in the tests and their economical scoring by machines which in turn guarantee the reliability issues of these discrete-point tests. However such assessments only attempt to test a few of the discrete subsets which constitute writing ability. These are typically the knowledge of particular linguistic features such as grammatical choices pertaining to sentence grammar, vocabulary or even spelling and punctuation (Cumming, 1997). Performance assessment, on the other hand, has become standard practice in L1 contexts and is also typically adopted in the context of L2 writing assessment. It requires test-takers to demonstrate not only language knowledge, but also performing that knowledge in communicative settings (Shohamy, 1983). However, such assessments are generally acknowledged to be more subjective and challenge the writing raters to utilise scoring criteria that can clearly articulate the construct of writing ability. As Bachman, Lynch, and Mason (1995) also noted, performance assessment brings with it "potential variability in tasks and rater judgements, as sources of measurement error" (p. 239). McNamara (1996) proposed a strong and a weak sense of performance assessment in an L2 writing context. In the former version, the test task will represent a real world task and performance will be judged on real world

criteria. Thus the primary emphasis is on the successful fulfilment of the task and not on the successful use of the language. In the latter, the main focus is on the display of different aspects of L2 writing ability. Table 2.2 summarises the prominent features of direct and indirect assessment (Hyland, 2003).

Table 2.2 Features of direct and indirect assessment (Hyland 2003, p. 217)

Indirect assessment	Direct assessment
Claimed objective measurement	Based on production of written texts
High statistical reliability	High validity
Allows standardisation	Reflects real-life communicative demands
Inferential judgement of ability	Ability judged directly
Easy to administer	Integrates all elements of writing
Easy to mark	Requires rater training

Previous research into the assessment of second language writing generally entails the direct assessment of writing performance, rather than indirect or objective tests of writing (Hyland, 2002). In a similar vein, Shaw and Weir (2007) argue that tests should be made as direct as practicable. They posit that generalising from these types of tests to how students may perform more productive tasks requiring construction of more complex texts is more plausible than from indirect tests which only measure parts of the construct of writing ability.

2.2.4.2. Types of rating scales

Once the assessment tasks are designed and administered and the written texts are collected, the issue of scoring them using a rating scale will be raised. A rating scale, which is sometimes referred to as scoring rubric or proficiency scale, is defined by Davies et al. (1999, p. 153) as:

> A scale for the description of language proficiency consisting of a series of constructed levels against which a language learner's performance is judged. Like a test, a proficiency (rating) scale provides an operational definition of a linguistic construct such as proficiency. Typically such scales range from zero mastery through to an end-point representing the well-educated native speaker. The levels or bands are commonly characterised in terms of what subjects can do with the

language (tasks and functions which can be performed) and their mastery of linguistic features (such as vocabulary, syntax, fluency and cohesion).

Hyland (2003) identified three scoring methods: 1) holistic, i.e. a single score of writing behaviour is given, 2) analytic, i.e. separate scales of overall writing features, each assessing a different aspect of writing are assigned, 3) trait-based scoring, i.e. the performance traits relative to a particular task are judged. The trait-based scoring method uses either primary-trait scoring, i.e. only specific writing features relevant to a task (e.g. appropriate text staging, creative response, effective argument, reference to sources, audience design) are rated, or multiple-trait scoring, i.e. separate scores for different writing features are provided by raters. What differentiates multiple-trait scoring from the analytic scoring is the flexibility of the scores adapted to the context, purpose, and genre of writing. The last two trait-based scoring methods have not been very common in L2 writing assessment, because devising and administering the rating scales take enormous amounts of time and they can only be used for one single task at a time. Holistic and analytic scales, however, have been most commonly used in L2 writing assessment. One example of high-stakes language tests commonly administered around the world is the International English Language Testing System (IELTS) which are scored based on a nine-band rating scale and are reported in a single score. The advantages and disadvantages of these two rating scales, as summarised by Weigle (2002), are presented in Table 2.3.

Table 2.3 Holistic vs. analytic rating scales (Weigle 2002, p. 121)

Quality	Holistic Scale	Analytic Scale
Reliability	Lower than analytic but still acceptable	Higher than holistic
Construct Validity	Holistic scale assumes that all relevant aspects of writing develop at the same rate and can thus be captured in a single score; holistic scores correlate with superficial aspects such as length and handwriting	Analytic scales more appropriate for L2 writers as different aspects of writing ability develop at different rates
Practicality Impact	Relatively fast and easy Single score may mask an uneven writing profile and may be misleading for placement	Time-consuming; expensive More sales provide useful diagnostic information for placement and/or instruction; more useful for rater training
Authenticity	White (1985) argues that reading holistically is a more natural process that reading analytically	Raters may read holistically and adjust analytic scores to match holistic impression

Summary

Research and theoretical perspectives on the processes involved in second language writing have evolved over the last forty years from one in which writing was seen as primarily a set of learned modes of discourse in five-paragraph essays in a product-centred pedagogy, to recognition of the importance of assisting writers to discover their own voices in a process pedagogy. Also, studies of academic literacy development in L1 and L2 have revealed that undergraduate writing is a set of practices which are socially contextualised rather than merely being the cause of the cognitive skills inculcated by given institutions (Grabe & Kaplan, 1996).

2.3. Planning and writing

A good number of studies have been conducted to investigate task-related factors which may have a bearing on fluency, complexity, and accuracy of task-based production. Among factors studied so far are task implementation variables including planning (Ellis, 1987; Ellis & Yuan, 2004; Mehnert, 1998; Skehan & Foster, 1997; Wigglesworth, 1997; Yuan & Ellis, 2003), and task design variables including type of input, task conditions, and task outcomes (Skehan & Foster, 1999; Swain & Lapkin, 2001). Planning, a task

implementation variable, is a problem solving activity which all language use, whether spoken or written, entails to help the language user to decide what and how to say/write. In essence, expert writers have been shown to often devote more than two-thirds of their writing time to planning which may occur before or during text production (Flower & Hayes, 1980; Gould, 1980). In contrast, it is probably not surprising to learn that even if promoted to do advanced planning, novice and struggling writers devote less than one-half minute, regardless of their age, the writing genre or the writing medium (De La Paz, 1999; De La Paz & Graham, 1997; Graham, 1990; Harris, Graham, & Mason, 2006; Lane et al., 2008; MacArthur & Graham, 1987).

As Clark and Clark (1977) noted, planning takes place at a number of different levels, resulting in discourse plans, sentence plans and constituent plans, all of which have to be interwoven in the actual execution of a language act. Flower and Hayes (1981), in their analysis of planning activity under think-aloud conditions, identified three distinct planning sub-processes: (1) idea generation, i.e. to retrieve relevant information from long-term memory, (2) organisation, i.e. to give a meaningful structure to the ideas, and (3) goal setting, i.e. to establish both procedural and substantive goals. They posit that throughout composing, writers' goals they have created can lead them to generate and organise ideas, and those ideas can in turn result in some newer goals. Based on this assertion, one may conclude that planning functions as a dynamic network which is not limited to the pre-writing stage, but instead is closely and inseparably prevalent in the ongoing and moment to moment composing process.

2.3.1. Types of planning

Burtis, Bereiter, Scardamalia, and Tetroe (1983) distinguished between two types of composition planning: conceptual planning (for the reader, purpose, goals, strategies, and organisation) and content planning (for generating information). In establishing a framework for understanding cognition in writing, Hayes (1996) hypothesised that planning enables leaners to allocate attentional resources between what they are going to produce (content planning) and how they are going to produce it (conceptual planning). In the extant literature of task-based planning (e.g. Ellis, 2005; Ellis & Yuan, 2004; Yuan & Ellis, 2003), two principal types of planning are distinguished, simply in terms of when the planning takes place, i.e. pre-task planning, planned before the task is performed, and

within-task planning, planned during the task performance, and can be differentiated according to the extent to which the task performance is pressured or unpressured.

Research on the effect of different levels of planning, e.g. pre-task planning and on-line planning (Ellis, 1987; Ellis & Yuan, 2004, 2005; Yuan & Ellis, 2003), the amount of time allocated for planning (Mehnert, 1998), the source of planning, and guided and unguided planning (Foster, 1996; Foster & Skehan, 1996; Ortega, 1999; Skehan & Foster, 1997) supports the role of planning in improving fluency and complexity of language production, whereas the effects on accuracy seems to be mixed. Some studies (e.g. Foster & Skehan, 1996; Mehnert, 1998; Skehan & Foster, 1997) do suggest that accuracy is raised when there is a kind of planning. For instance, Ellis and Yuan (2004) found that online planning improved accuracy and grammatical complexity, but other studies (e.g. Crookes, 1989; Ortega, 1999) do not support this claim.

2.3.2. Theoretical frameworks

All these studies investigating the impact of planning on language production, either implicitly or explicitly, draw on information processing theory which claims that human beings possess a limited processing capacity and thus are not able to attend fully to all aspects of a task (Newell & Simon, 1972; Robinson, 1995). This is apparently far more difficult for L2 learners, particularly those with limited proficiency, to attend to both form and meaning simultaneously and thus have to make decisions about how to allocate their attentional resources by prioritising one aspect of language over others (Robinson, 1995; Skehan, 1998; Skehan & Foster, 2001). However, according to planning research, when learners are given the opportunity to plan the linguistic and propositional contents of an upcoming task, they can better compensate for their processing limitations, and thus their linguistic output quality is enhanced significantly (Skehan, 1996). In what follows, two central constructs which have been involved in cognitive and psycholinguistic accounts of language processing and are bound up with the study of task planning will be presented. These are planning and working memory capacity, and planning and the cognitive models of task-based performance. In each case, I will outline the underlying theories and consider how it may be applied to task planning in writing.

2.3.2.1. Planning and working memory

Baddeley (1986) defined working memory as "the temporary storage of information that is being processed in any range of cognitive tasks" (p. 43). One of the most frequently cited models of working memory in task planning literature belongs to Baddeley. This tripartite model is composed of three main components: the central executive which functions as supervisory attentional system, the phonological loop which stores phonological information and prevents it decay, and the visual spatial sketchpad which can be divided into separate visual, spatial and kinaesthetic components (see Figure 2.2).

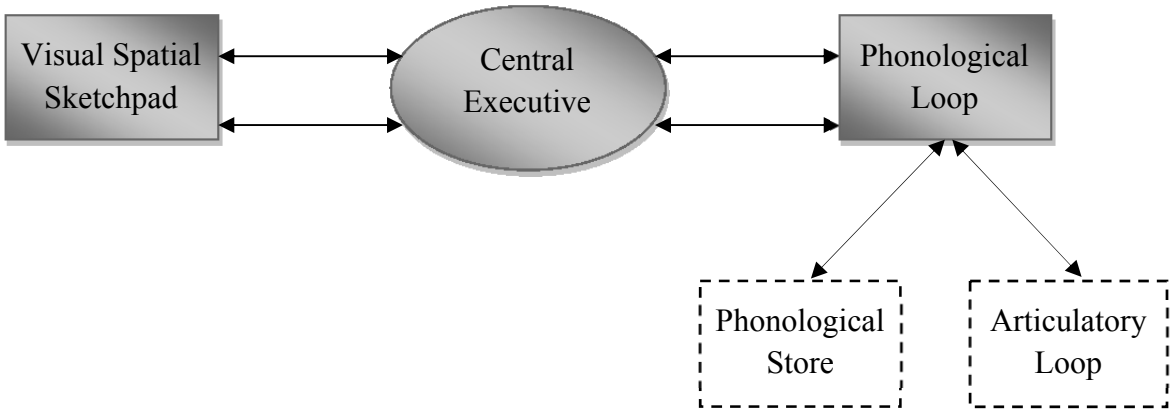


Figure 2.2 Baddeley's model of working memory (1986)

According to Ellis (2005), only the first two components are pertinent to task planning and not the visual spatial sketchpad. The central executive system is generally considered to have limited capacity; however, the extent to which other systems are automatised can direct attention to relevant information by coordinating cognitive processes when more than one task needs to be done at the same time. Strategic activities like preplanning, making notes, outlining, and so forth are likely to function as a facilitative role in lessening an overloading of the working memory. The phonological loop consists of two sub-components: a short-term phonological store which is subject to rapid decay and an articulatory loop which can revive the memory traces by continuously articulating its contents. By the same token, providing learners with the opportunity for planning allows learners to access and maintain one set of linguistic material in the phonological loop while they draw on another set of material to monitor, modify or refine it (Baddeley & Logie, 1999; Kellogg, 1990; Skehan & Foster, 1999).

In his model of writing processes which was presented in section 2.2.2 Models of L1 and L2 writing process, Kellogg (1996) expanded the earlier models of writing (e.g. Baddeley & Hitch, 1974; Flower & Hayes, 1980) and related the three basic processes involved in writing (i.e. formulation, execution, and monitoring) to Baddeley's (1986) model of working memory.

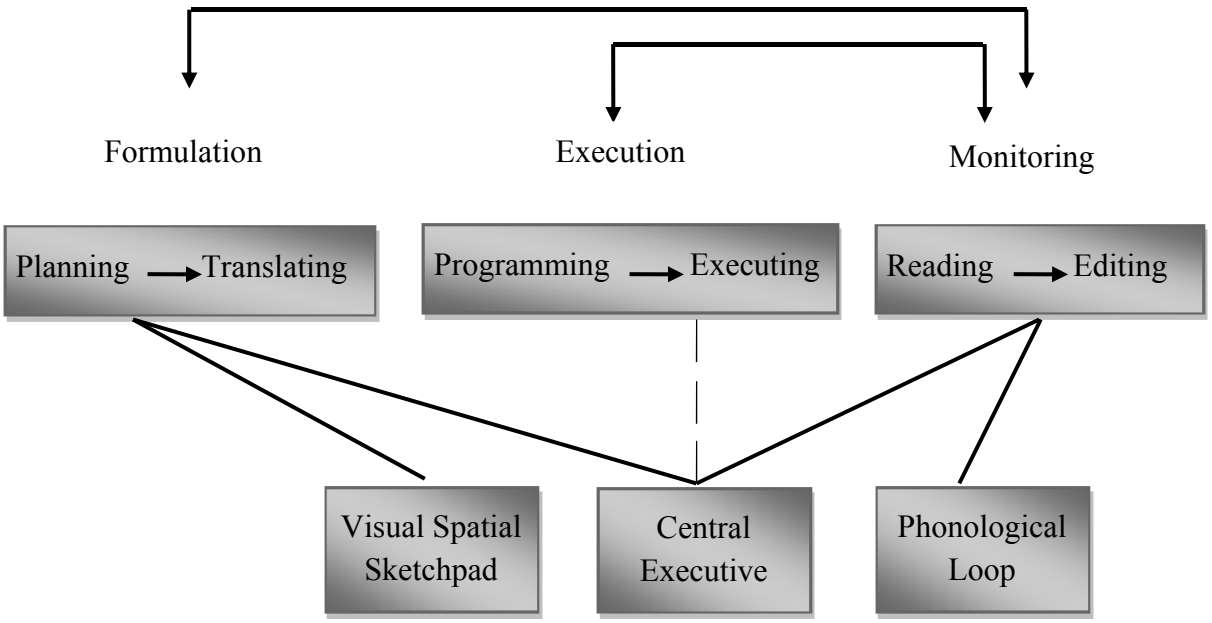


Figure 2.3 Kellogg's model of writing processes (1996)

As diagrammatically demonstrated in Figure 2.3 above, all the sub-processes except executing involve the central executive. Ellis and Yuan (2004), however, argue that this may be true for an adult native-like writer who does not require controlled processing for handwriting or typing. For L2 learners, particularly those with different scripts in L1, the central executive is likely to be called upon. In his model, Kellogg also proposed that the visual spatial sketchpad is only involved in planning and the phonological loop is required for both reading and editing. This is partially in contrast with Ellis (2005) who posited that the visual spatial sketchpad does not seem to be relevant to the role of task planning.

2.3.2.2. Planning and cognitive models of task-based performance

In this section, two influential albeit contradictory models of task complexity in task-based performance and learning will be compared and their relations with task planning will be discussed. These models are Skehan and Foster's (2001) Limited Attentional Capacity Model, and Robinson's (2001a, 2001b, 2003) Cognition Hypothesis which have been proposed to explain the allocation of attentional resources to different aspects of language production during task completion.

Skehan and Foster's (2001) Limited Attentional Capacity Model assumes that learners' attention and memory are limited in capacity such that the trade-off effects among the three aspects of language production, i.e. accuracy, complexity and fluency, become essential. This is because, they claim, these aspects draw on different systems of language. Based on Skehan's (1998) Cognitive Approach, there are two distinct language systems: exemplar-based and rule-based. The former contains linguistic knowledge related to discrete lexical items and ready-made formulaic chunks of language, whereas the latter includes abstract representations of the underlying patterns of language. Skehan argues that the exemplar-based system can be easily accessed and thus is best suited for fluent language performance, while the rule-based system requires more processing and thus is ideally suited for more complex or accurate language performance as a result of increased input corporation and output modification. What distinguishes complexity from the accuracy is the extent to which learners take risks to access and restructure the cutting edge interlanguage features (as in the case of complexity) or the extent to which learners attempt to access more fully acquired features and avoid errors in order to perform according to target language norms (as in the case of accuracy). In their model, Skehan and Foster (2001) posit that increasing cognitive task complexity will divert learner's attention from the complexity and accuracy of their language production to the task content development.

In Robinson's (2001a, 2001b, 2003) Cognition Hypothesis, cognitive task complexity is determined by two sets of dimensions: resource directing and resource depleting features, which are thought to "interact and affect task production in measurable ways" (2001b, p. 31). The resource directing features of the task make conceptual demands on learners (e.g. from Here and Now tasks to There and Then tasks, from whether tasks require reasoning

to tasks which do not, from tasks which require learners to make references to few elements to tasks with many elements). The resource depleting features of task performance, on the other hand, make procedural demands on learners (e.g. from planning to no planning, from single tasks to dual tasks, from tasks which are provided with prior knowledge to tasks which are not). Contrary to the Limited Attentional Capacity Model, Cognition hypothesis postulates that leaners have multiple attentional resources. Robinson predicts that increasing task complexity along the manipulation of resource-directing variables will lead to greater increased accuracy and complexity of production, whereas an increase in task complexity with respect to resource-depleting features can facilitate the development of more fluent performance, but will affect accuracy and complexity negatively.

It can be clearly concluded that these two models afford contrasting predictions as to the effects of planning on language performance, which is also relevant in studying writing performance. In Skehan and Foster's model, strategic planning helps learners to access their rule-based, rather than their exemplar-based or memory-based, system. Thus, they speculate that strategic planning is likely to enhance linguistic complexity but not accuracy. In contrast, planning is incorporated as a resource depleting factor in Robinson's Cognition Hypothesis and thus results in increased fluency to the detriment of decreased accuracy and complexity.

2.3.3. Empirical studies of planning and writing task performance

Unlike the large number of studies examining the effects of task planning on oral performance, there have been comparatively fewer empirical studies investigating the impact of planning on written performance. In both oral and written task-based language teaching research, measures of fluency, complexity, and accuracy are commonly used. Fluency refers to "the production of language in real time without undue pausing or hesitation" (Ellis & Barkhuizen, 2005, p. 139), complexity is "the extent to which learners produce elaborated language" (ibid, p. 139), and accuracy refers to "how well the target language is produced in relation to the rule system of the target language" (Skehan, 1996, p. 23). Furthermore, although M. D. Johnson, Mercado, and Acevedo (2012) posit that the predictions of the two models of Robinson's Cognition Hypothesis and Skehan and Foster's Limited Attentional Capacity model about L2 oral language production do not

apply to the manipulation of cognitive processes on writing quality, these two models are extensively used in both oral and written production research. In the following section, the extant research studies on task planning in writing will be reviewed in a chronological order, grouping, as far as is possible, those studies which have been somewhat similarly designed with regard to task planning. This review of the planning research in writing will include studies in L1 and L2 which have investigated different variables in terms of planning conditions, writing task types, and the aspects of writing performances under analysis.

The effects of planning sub-processes have been investigated more in L1 writing than in L2. Idea organisation and goal setting in combination have been found to have positive effects on the quality and fluency of L1 writing (De La Paz & Graham, 2002; Graham & Perin, 2007; Page-Voth & Graham, 1999; Sexton, Harris, & Graham, 1998; Troia & Graham, 2002). In L1 writing research, Kellogg (1987) found that organisation pre-task planning in the form of outlining positively impacted fluency and grammatical complexity of texts. With respect to L1 college students' persuasive writing, Kellogg (1988) examined the effects of outlining, no outlining, and mental outlining on the fluency and text quality. The outlining group was instructed to work on an outline for 5-10 minutes before writing, the no outlining group was asked to write immediately, and the mental outline group was instructed to create a mental outline, rather than a written plan, of what they would write. The results indicated that preparing an outline, whether written or mental, did not improve fluency but it led to higher quality written texts. As a replication of his results on outlining, Kellogg (1990) examined L1 college students' argumentative writing quality under three planning conditions (no pre-writing, outlining, clustering) and three sub-planning conditions (topic given, topic and ideas given, and topic, ideas and organisation given). The pre-writing condition was instructed to begin drafting immediately, the outlining condition was asked to prepare a hierarchical written outline, and the clustering condition was instructed to plan by linking related ideas using a visual network. The results showed that outlining had a significant effect on the fluency of text production and the overall text quality, particularly for the topic condition, but not for the topic, ideas, and organisation condition. The number of ideas generated during prewriting was shown to have increased in the clustering condition, but this had no impact on the quality and fluency of the texts. Only when fluency was measured by the total time on task, and not

the amount of writing time, it was highest in the control group who had begun drafting without any prewriting time.

Ellis (1987) compared the L2 learners' performance on written and oral pictorial narrative tasks under three conditions: planned writing, planned speech, and unplanned speech. As regards the written task, the learners were allowed to write the narrative in their own time. The findings revealed that when there was no time pressure thereby allowing for online planning, the accuracy of learner's use of regular past tense forms was significantly higher in the written narratives than in the oral task which was retelling the same narrative but without recourse to the written versions. Building on this study of Ellis (1987), and Yuan and Ellis (2003), which compared the effects of pre-task and online planning on L2 learner's monologic oral narrative production, Ellis and Yuan (2004) conducted a similar study but this time examined the effects of different types of planning on L2 narrative writing performance. The results in this study did not completely corroborate those obtained in Yuan and Ellis (2003). The findings demonstrated that the pre-task and online planning impacted different aspects of L2 writing processing. They discussed that pre-task planning promoted the formulation process of the task and thus resulted in greater fluency, and syntactic complexity and variety. Online planning, on the other hand, provided better opportunities for monitoring, and thus assisted greater error-free clauses. Pre-task planning was shown to have less effect on accuracy, and online-planning resulted in decreased fluency, but fewer dysfluencies, a difference which showed a tendency to statistical significance. In comparison to the two planning groups, the no planning condition resulted in negative consequences for the three measures of fluency, complexity, and accuracy of the written product. Ellis and Yuan (2004) asserted that this was because they were required to formulate, execute, and monitor under pressure.

In examining the effects of planning in argumentative composition, Dellerman, Coirier, and Marchand (1996) asked the L1 proficient and non-proficient writers in secondary school to complete a constrained argumentative writing in 30 minutes, based on the 13 arguments provided to them. The results did not show any global effect of planning on the quality of written texts. They revealed, however, that planning focusing on logical relationships had a significant effect on the argumentative texts particularly for non-proficient writers, based on the assumption that planning would improve the organisation of information and increase the available cognitive resources for high-level processes.

In another study, Shi (1998) investigated the effects of prewriting discussions on the quality of ESL pre-university students' opinion essays. She studied prewriting under three different conditions: peer-talk discussion, teacher-led discussion, and no discussion. Shi found that students wrote longer drafts in the no discussion condition, shorter drafts in the teacher-led discussion group, and more complex drafts with a wider variety of verbs in the peer discussion condition. This study is one of the seminal studies which investigated prewriting planning within the sociocultural dynamics of peer or group responses to writing. One limitation, however that some researchers such as Shin (2008) identify in Shi's study is that if she incorporated an individually planned condition rather than the no discussion condition in her study, it could have yielded more comparable results about task planning.

Ojima (2006) used a case study to examine the effect of concept mapping strategy as a form of pre-writing planning on three Japanese ESL students' written performance. Each learner was instructed to write four compositions: the first two were unplanned (one written in class, and one at home), the next two were pre-planned after the concept mapping strategy was introduced to them (one written after individual planning in class, and one at home after having brainstormed during group discussions in class). The composition data were analysed both by holistic measures and by accuracy, complexity, and fluency measures. The results showed that drawing concept mapping for in-class compositions produced greater complexity and fluency, but did not improve accuracy. The holistic analysis, however, showed no significant differences between the pre-planned and non-planned compositions. Ojima reported that these findings corroborated Ellis's (2003) summary of strategic planning features, which he had reviewed from the empirical studies in planning research. In that review, strategic planning was shown to have a significant effect on complexity and accuracy, but only on those grammatical features of accuracy which were rule-based. Having compared the effectiveness of individual and group planning, Ojima implied that individual planning is likely to be more effectively applied to the students' composing processes than group planning because, she asserted, concept mapping and drafting activities directly function at conceptual and textual levels of written production. This case study, however, lacks the generalisability power due to the small sample size.

Shin (2008) investigated the effects of individual planning and collaborative planning on Korean learners' expository and argumentative writing productions. Over a period of two weeks, the learners in the individually planned condition were asked to plan for 10 minutes on their own before writing an essay for 30 minutes, whereas the learners in the collaboratively planned condition were allowed to plan their writing with a peer but were required to independently complete the essays. The results indicated that collaborative planning had a significant impact not only on the overall quality of expository writing tasks but also on all their five analytic measures covering the areas of content, organisation, language in use, grammar, and mechanics. Shin, however, found no significant differences between the two types of planning conditions regarding the overall quality or either of the analytic features in argumentative writing tasks. In short, peer collaboration appeared to have been more effective in expository writing tasks than in the argumentative writing tasks. Shin explained that this could be due to the fact that learners managed to brainstorm their ideas better and build up more reasonable explanations in the expository writing task than in the argumentative writing tasks which are perceived to be more difficult to fulfil in educational contexts. However, the extent to which these reasons are compelling is untenable. It is not clear how Shin assessed learners had a better opportunity to brainstorm in expository tasks. Moreover, she did not empirically investigate the degree of complexity or difficulty of the two types of tasks to claim that argumentative tasks were more cognitively demanding.

In a quasi-experimental study, Dujsik (2008) explored the effects of pre-writing strategy training guided by Inspiration 6, an idea graphic organiser software program, on ESL students' writing quality and quantity. Forty one intermediate participants were assigned to two control and two experimental groups. The instructional modules for the groups differed: the control groups were taught paragraph writing, essay writing, and opinion essays, whereas the experimental groups were instructed pre-writing strategies related to writing purposes, audience, and idea generation and organisation. The pre- and post-test comparisons demonstrated that there were no significant differences between the groups in terms of the quantity and quality of writing production, however, the strategy-trained groups showed a tendency to improve in their writing quality variables.

Pu (2009) examined the language quality of 24 Chinese first-year English majors' argumentative writing tasks. There were four groups in the study: one group was

instructed to individually brainstorm prior to the writing task, the other three groups were differentiated based on the 15-minute prewriting discussions they were required to have in either English, Chinese, or both English and Chinese. The results of the study revealed that the three prewriting discussion groups wrote more fluently than the brainstorming group. The between-group comparisons for the discussing groups showed that the group discussing in English performed better than the Chinese group, and the English/Chinese group in terms of the fewer number of grammatical errors and higher syntactic complexity. In another study similar to the investigations on the effects of prewriting discussions in different languages, Karim (2010) investigated the effectiveness of prewriting discussions in EFL writing through comparing the use of L1 and L2 in discussions. Thirty sophomore EFL students were instructed to write four essays after prewriting discussions: two essays after discussions in their first language (Kurdish), and two essays after L2 discussions. The findings showed that the L2 prewriting discussions were more effective than the L1 discussions in students' overall text quality using Cohen's (1994) analytic scoring rubric. The results of these two studies contrasted with those obtained in Lally (2000), and Stapa and Abdul Majid's (2009) studies, which showed that prewriting activities in L1 were more effective in students' writing performance.

Ong and Zhang (2010) attempted to systematically examine the combined effects of resource directing and resource dispersing variables on writing performance, according to Robinson's (2001a, 2001b, 2003) Cognition Hypothesis. In their study, the effects of task complexity on the fluency and lexical complexity of Chinese EFL learners' argumentative writing tasks were manipulated using three factors of planning time, provision of ideas and macro-structure, and the availability of first drafts during revision. The findings of the study revealed that increasing the planning time continuum produced greater fluency and lexical complexity. Thus, they concluded that these findings lent support to the predictions of neither Robinson's Cognition Hypothesis, nor Skehan and Foster's (2001) Limited Attentional Capacity. Further, the provision of ideas and macro-structure resulted in significantly greater lexical complexity but not fluency. This, they posited, only partially supported the predictions of the two above-mentioned models. Finally, the task complexity manipulated with respect to draft availability had no significant effects on fluency and lexical complexity in the revised drafts. This result again did not support Robinson's Cognition Hypothesis.

In a study in which pre-task planning was coupled with the immediacy of time and space, Khomeijani Farahani and Meraji (2011) examined the effects of these two task design features on the complexity, fluency, and accuracy of Iranian EFL students' written narrative performance. The participants were assigned to one of the four groups, depending on two levels of pre-task planning and no pre-task planning conditions, and two levels of +/– Here and Now conditions. As regards accuracy, the results showed that the two planning conditions outperformed the two no planning conditions, however the Here and Now conditions and the There and Then conditions performed equally on the accuracy measures. In terms of syntactic complexity, the researchers showed that the students' writing improved when provided with time to plan, but not with increments in task complexity along immediacy per se, however the interaction between these two dimensions of planning and immediacy significantly influenced syntactic complexity. Concerning the lexical complexity measures, no differences were found in the planning vs. no planning, in Here and Now vs. There and Then groups, or in the interaction of the two. Finally, the provision of planning time and the manipulation of immediacy resulted in significantly more fluent text production in terms of length of texts and the number of dysfluencies. They concluded that their study lent support to the predictions of Skehan and Foster's (2001) Limited Attentional Capacity model, as accuracy could not keep pace with the higher syntactic complexity measures due to leaners' limited attentional resources although they reported effects for fluency were not deteriorated.

In another study again conducted in the Iranian EFL context, Mahnam and Nejadansari (2012) revealed the positive effects of pre-writing strategies on students' writing achievement. They investigated the effects of different pre-writing strategies, namely concept map, reading relevant texts, and negotiation, on the quality of argumentative writing essays. This research successfully demonstrated the effectiveness of teaching planning strategies to students in EFL settings. It may be argued, however, that the participants in this study were all advanced-level adult students and thus the explicit instruction of pre-writing strategies would be more effective and could bring about marked differences in the post-test writing achievement of the experimental group.

Only M. D. Johnson, Mercado, and Acevedo (2012) have investigated the specific impact of planning sub-processes on L2 writing. The remaining L2 research has identified the effects of planning on the quality of texts in a general sense and not of each sub-process.

Johnson et al. found that pre-task planning had no consequential impact on lexical and grammatical complexity of L2 writers. They suggested three explanations to account for their findings: (1) Speaking is a linear process, whereas writing is a recursive process. Thus the predictions of the two models of Robinson's (2001a, 2001b, 2003) Cognition Hypothesis and Skehan and Foster's (2001) Limited Attentional Capacity model about L2 oral language production do not apply to the manipulation of cognitive processes on writing quality. (2) The impact of pre-task planning on L2 writing may be moderated by the learners' educational background, educational experience, and genre familiarity. In Johnson et al.'s study, the participants had received instruction on the composition of comparison-contrast essays during their writing courses, however they were assigned argumentative prompts to write to, that is to say they were called to compose in an unfamiliar genre. (3) In order to be freed from the demands of the working memory resources, the proficiency of the participants need to achieve the proficiency threshold. In this study, however, the writing proficiency of the majority of participants was limited.

Haghverdi, Biria, and Khalaji (2013) investigated the effects of task planning and gender on the accuracy of Iranian EFL learners' pictorial narrative writing. The three levels of planning operationalisation followed those of Ellis and Yuan (2004). Accordingly, the task planning implementation differed, based on the groups having been assigned to three conditions: no planning, strategic planning, and within-task planning. The results showed that when provided with 10 minutes to plan for their narrations prior to performing the task, the participants outperformed in their production of texts which contained accurately used verbs and clauses. As far as gender was considered, they found that gender did not have any significant effects on the mean differences of the three groups. The researchers concluded that strategic planning was likely to be more effective than within-task planning in Iranian contexts. This is an issue worth further investigations in future task-based research in either Iranian, or non-Iranian settings.

Summary

Although task planning has dominated in task-based language research, significant discrepancies have been reported in L2 cross-sectional studies with regard to the effects of planning on the quantity and quality of individual writing tasks. The need for planning studies of how learners perform a series of tasks over a substantial period of time has been noted (Ellis, 2003, p. 137). The former studies provide information about how language is

used to foster accuracy, complexity, and fluency, whereas the latter research sheds light on the acquisition of language over time.

Conclusion

The robust research identified throughout this chapter informed the overarching aim of this study and provided the basis of the design and planning of the writing and metacognitive intervention syllabi. First, research into the theoretical underpinnings of different models and approaches of writing were discussed to present useful vantage points for the current study on the whole. Secondly, a good number of studies which have identified the impact of strategy instruction on writing production were reported. In particular, research undergirded by the metacognitive instruction in writing was reviewed. Finally, the impact of planning on writing accuracy, complexity and fluency which have been examined in several task-based writing performance studies were presented. Working memory was shown to have an integral role in the writing process, particularly in the planning and production of written texts. More specifically, it was discussed that the examination of factors affecting planning processes has recently gained more attention in L2 writing teaching and learning and studies of this type have proved to be beneficial in that they provide us with a better insight of how planning functions.

There remains a question as to how the three constructs are interrelated. Although the relationships among the three constructs are theoretically well-grounded, they are still in need of further empirical investigation. To the best of my literature review search, no previous study has to this date investigated the interrelationships among the three constructs. Accordingly, the present study built on the existent literature by incorporating the following elements: detailed analyses of written texts, for explaining the differences and/or similarities in the writing practices of groups receiving different writing and/or metacognitive strategy interventions, and the think-aloud protocol and questionnaire analyses, for reaching an in-depth understanding of the composition process. The former was informed by the theoretical principles of text-oriented approach to writing, while the latter was informed by the principles of writer-oriented and reader-oriented approaches to writing. To this end, it will be argued in this study that the traditional methods of teaching writing emphasising on the mechanical micro-level aspects of writing should give way to more effective metacognitive strategy instruction which allows students to focus on pre-writing, planning, organising, monitoring, and revising.

CHAPTER 3. RESEARCH METHODOLOGY

The purpose of the present chapter is to give a detailed account of the research design of the project and to outline the principles underpinning the choice of the research methodology. An essential element in designing a study is the use of data collection instruments and data analysis categories which have been carefully trialled. In this project, Study 1 functioned as a linked study undertaken initially to enable me as the researcher to evaluate and confirm the research procedures to follow in Study 2. This process resulted in some adaptations to data collection, data organisation and data analysis.

In the first part of this chapter, the research design and the research questions are presented. The setting and sampling strategies are then outlined. This is followed by a description of data collection instruments and procedures. The rationale and procedure for data analysis are then described. The final section of the chapter discusses strategies used to enhance the quality and ethical integrity of the project.

3.1. Research approach

The current study aimed to find answers to the following research questions:

1. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *general writing proficiency* of L2 learners' written products?
2. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *accuracy* of L2 learners' written products?
3. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *complexity* of L2 learners' written products?
4. What metacognitive strategies do L2 learners perceive in their actual writing performance?

Thus, the study involved an empirical study with four conditions. Overall, it was a double-factor, between-groups design with two levels of planning condition (pre-task planning and no pre-task planning) and two levels of metacognitive training condition (trained and

non-trained groups) on argumentative writing tasks. According to Ellis and Barkhuizen (2005), studies conducted in social sciences are illustrative of three research paradigms; namely, normative, interpretative and critical. In terms of theoretical orientation, research purpose, and design, the present study mainly draws on the normative paradigm. It serves to establish whether learners, in EFL or ESL contexts, can manifest a better performance on argumentative writing tasks when their metacognitive awareness is raised. To test this hypothesis, an experiment involving randomly constituted groups of L2 learners and a carefully contrived treatment was conducted. The study utilised both normative and interpretative approaches, in terms of methods of data collection and data analysis. In effect, in addition to the quantitative methods which were supported by inferential statistics to identify cause and effect relationship between independent and dependent variables and significant differences among groups, qualitative methods involving verbalisation recordings of participants during writing were also utilised.

A mixed-methods methodology was undertaken in the design of this research. Teddlie and Tashakkori (2009) maintain that at least one quantitative and one qualitative method is used in the mixed methods study either at the data collection or data analysis of the research, "where neither type of method is inherently linked to a particular inquiry paradigm or philosophy" (p. 323). In that sense, they argue that the mixed methods researchers aim at a workable middle position from multiple viewpoints. Inherent in the use of such methodology is therefore the triangulation of the data derived from the corroboration of quantitative and qualitative findings in one single study (Dörnyei, 2007). Collating data from a range of sources which are collected through a range of research methods can "map out", or explain more fully, the "richness and complexity" of a study (L. Cohen, Manion, & Morrison, 2000, p. 112). In recent years, the use of mixed methods has also become increasingly common in studies in the field of second language acquisition and applied linguistics. Still, over the 1995-2005 period, 6.8 per cent of research papers in Applied Linguistics have been shown to subscribe to mixed-methods research engaging a multiple set of paradigms (Angouri, 2010). As discussed in Chapter 2, Dujsik (2008) is one example that used both quantitative and qualitative methods to investigate the effects of pre-writing strategy training on ESL students' strategy use, and writing quality and quantity.

3.2. Study 1 and 2

The project consisted of two studies (see Figure 3.1). The first study, which is Study 1 throughout this thesis, was conducted at a university level in an ESL context for six weeks in July and August 2010, with the aim of exploring the research questions and identifying any potential threats to the validity of the research design and instruments. Study 2 took place at a language school in an EFL context for nine weeks between April and June 2011, involving a group of participants at a level of proficiency similar to the ones in Study 2.

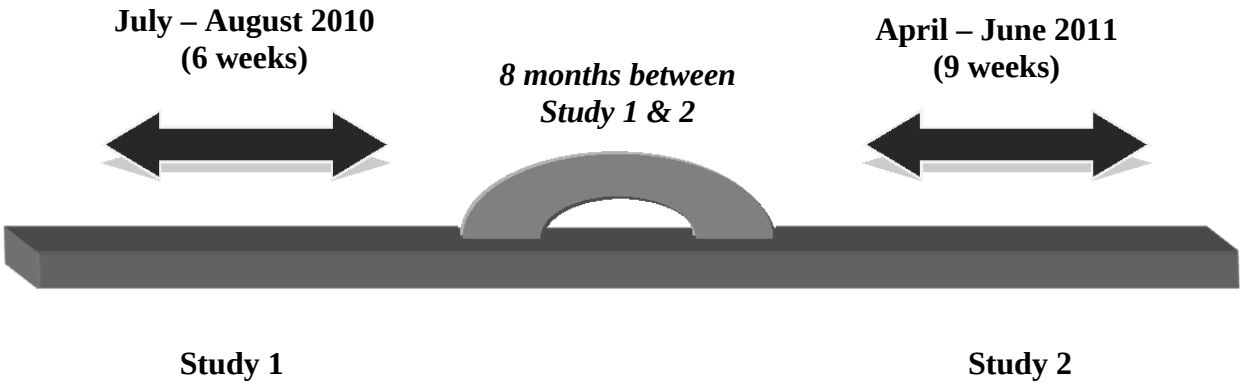


Figure 3.1 Timeline for data collection

The specific purposes of conducting Study 1 were to:

- assess the feasibility and testing the integrity of the procedure for data collection;
- trial the suitability of the materials, instruments, writing topics, writing models and samples to be used in the study;
- measure the timing and division of sections in the lesson plan in each session;
- justify the research questions in the real research context;
- ascertain whether think-aloud protocols and questionnaires would be effective instruments to help provide sufficiently rich process-enquiry data;
- develop the initial coding scheme of the qualitative data.

In this project, the purpose of using more than one approach to investigate the research questions was to maximise the credibility of the study and the internal validity of its findings. Out of the four types of triangulation, i.e. data, investigator, theory, method (Denzin, 1970), this study primarily drew upon two types: methodological triangulation (at the level of design and data collection), and data triangulation (involving time, place and

person). As illustrated in Figure 3.2, two different qualitative and quantitative methods were employed to collect data. Information was obtained from a range of sources (written texts, Likert-scale questionnaires, and verbalisation recordings of volunteer participants) in two different contexts and during two different time periods.

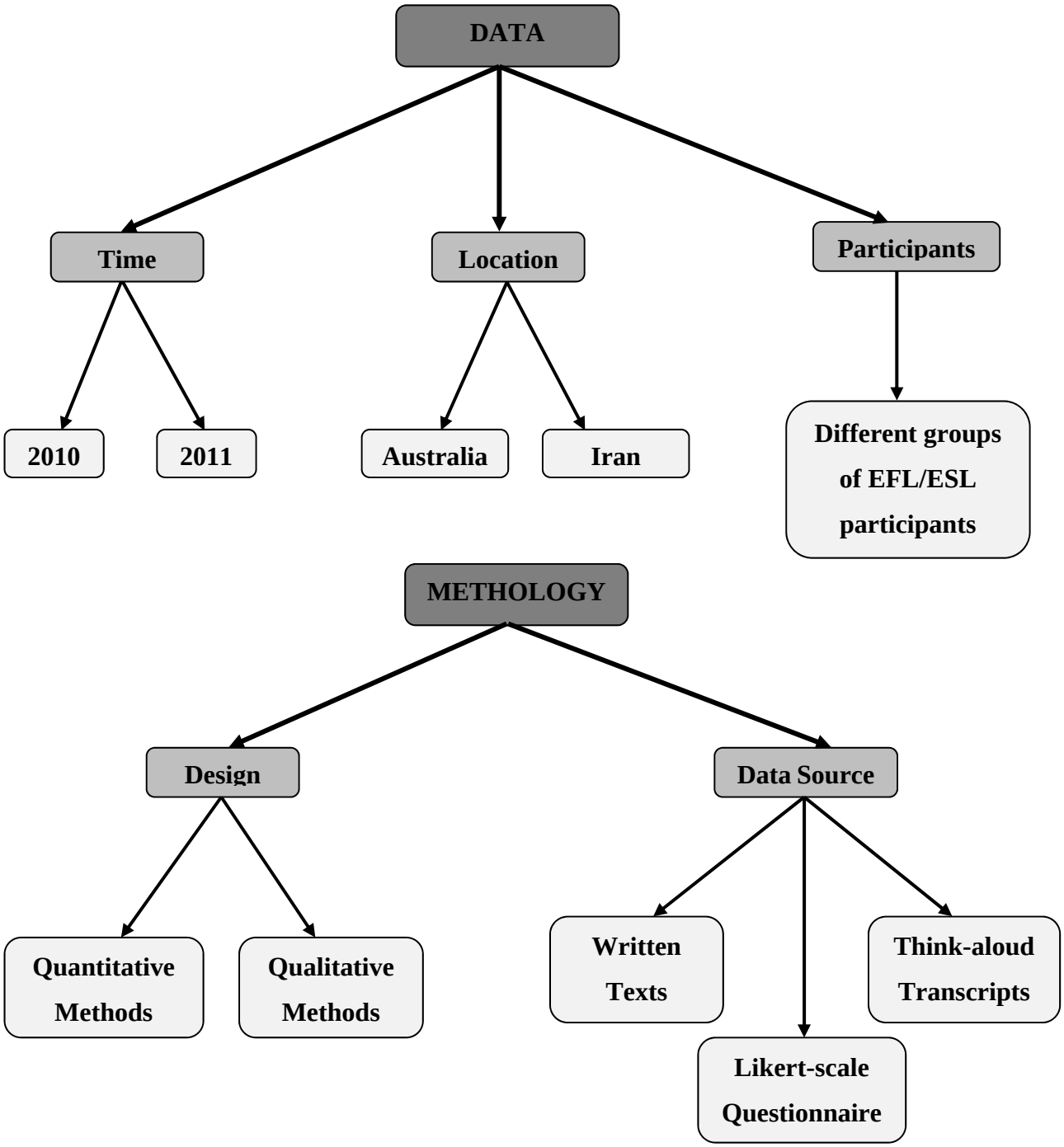


Figure 3.2 Types of triangulation in current study

3.3. Context and Participants

This section provides information about the characteristics of participants and the context from which they were recruited for the two studies of this project.

3.3.1. Context

Participants in Study 1 were undergraduate students enrolled in an academic program of English as an International Language (EIL) courses at a university in Melbourne, Australia. They had all satisfied the minimum English language requirements for international students from non-English speaking background to gain admission to undergraduate programs at this university, i.e. either a) English as the language of instruction in previous years of study; b) a minimum overall score on the Academic IELTS examination of 6.5, with no individual band score below 6, or an equivalent score on TOEFL exam (paper-based test: a minimum test score of 580 with a Test of Written English (TWE) score of at least 4.5; or internet-based: a minimum test score of 90 overall, at least 22 in writing and no section lower than 20); or c) successful completion of the English language bridging program offered by this university. At the time of data collection, participants were enrolled in subjects ranging from first to third year levels, and they had already completed the first two weeks of their courses. The first year students studying a sequence in English as an international language needed to complete two units (12 points). Second or third year students studying a minor or major in English as an international language had to have completed the first-year sequence. In addition, a minor required the completion of a further two units (12 points), and a major required the completion of a further six units (36 points). A minimum of three units needed to be taken at third year level. The range of subjects offered within the program included both a practical and theoretical understanding of the use of English for international communication. This content-based instructional program was intended to assist students who wished to widen their knowledge of the use of English for local as well as international purposes. More importantly, this program encouraged students' personal language development while not directly focusing on the explicit instruction of techniques and strategies of reading, writing, speaking and listening skills.

Participants in Study 2 were Iranian students of English as a Foreign Language (EFL) in a language institute in Mashhad, Iran. English education in Iran is introduced from the

beginning of junior high school which is a period of three years and covers grades six to eight. Overall, the mean hours allocated for English instruction at schools in Iran is about 450. This means that all participants had studied English for two hours per week for first grade of junior high school, and four hours for the second and third grade junior high school. There are three hours per week exposure to English teaching for the first grade high school, two hours of instruction per week for second and third grade, and four hours of instruction per week for the one year of pre-university. In order to obtain a better English proficiency, two thirds of students had attended English language institutes, on an average of three years, in addition to their formal educational syllabus at schools. Neither junior high schools nor high schools in Iran concentrate upon helping students practice and improve writing skills, and thus essay writing is not part of the curriculum. It is assumed that university students will develop writing skills on their own, after they have been provided with a thorough knowledge of lexico-grammatical rules and practice in translation. The main objective of this translation-oriented system is to enable students to take the National University Entrance Exam and to read and understand materials written in English as part of their disciplinary studies. The assessment system of English at junior high and high schools is based on written exams which contain sections on spelling, vocabulary, grammar and reading comprehension. Oral skills are practiced and tested via rote memorisation of dialogues and short conversations presented in textbooks. Pronunciation and intonation are assessed by requiring students to read aloud from texts.

3.3.2. Participants

As shown in Table 3.1, nearly a quarter of the 35 participating students in Study 1 were male, and just over three-quarters were female. There was no hypothesis, however, that gender would play a part in the analysis of the results. They ranged in age from 19 to 33, with a mean age of 21.74. They all spoke English as their second language. They came from seven different countries (China, Hong Kong, South Korea, Japan, Malaysia, Singapore, and New Caledonia). They were identified by the EIL program as being at an Upper Intermediate proficiency level.

As in Study 1, the 70 participating students in Study 2 were not controlled for gender, since this was not a variable in the design of this study. Gender wise, 38.57% of the participants were male and the rest were female. They ranged in age from 18 to 32, with a

mean age of 22.5. They all spoke Persian as their mother tongue. Only three (4.2%) had previously been taught by a native speaker of English, and around 34 (48.5%) had travelled overseas briefly as tourists. They were selected on the basis of their test scores from among a group of 100 candidates who had enrolled to take IELTS classes at this language institute which prepared students for taking the General or Academic IELTS examinations.

Table 3.1 Descriptive data of participants' characteristics

			Age		Gender	
	Groups	**Number**	**Minimum**	**Maximum**	**Male**	**Female**
Study 1	**+M+P**	10	19	22	1	9
	+M-P	10	19	24	2	8
	-M+P	8	19	33	3	5
	-M-P	7	20	29	2	5
	Total	35	M = 21.74		Valid Percent = 22.86	Valid Percent = 77.14
Study 2	**+M+P**	14	18	30	6	8
	+M-P	13	19	28	4	9
	-M+P	14	18	32	5	9
	-M-P	14	20	31	4	10
	Control	15	19	27	8	7
	Total	70	M = 22.5		Valid Percent = 38.57	Valid Percent = 61.43

Note. M = Metacognitive instruction, P = Pre-task planning

In Study 1, the researcher sent an email through the course coordinator requesting voluntary participation in the project to undergraduate students in the department. Subsequently, the researcher attended the first five minutes of a lecture in students' other courses and explained further about the workshop which was entitled "Improving Effective Academic Writing in English". During this short introductory spiel, the researcher answered students' questions about the project. Potential participants then contacted the researcher either by filling out the recruitment form or via email.

In Study 2, a notice was put up on the bulletin board of the institute for the voluntary recruitment of participants and an advertisement was posted on the official website of the

institute. Potential participants were those students who were applying to take IELTS preparatory courses and thus were initially required to take the institute placement test. They were then encouraged to attend my workshop on a voluntary basis. In order to explain further about the experiment, I attended the first five minute of their placement test session which was conducted on a monthly basis in this institute. Potential participants then contacted the institute coordinator to announce their voluntary participation.

Summary

This section provided a review of the background and context of the study. It also explained the rationale for selection of participants and the random sampling procedures which were carried out in order to secure 35 participants in Study 1 and 70 participants in Study 2. Explanations for the choice of instruments and the data collection procedures are presented in the following section.

3.4. Instrumentation and procedures

In this project, a variety of instruments were used to secure an in-depth understanding of the dynamic nature of the treatment. To ensure that participants were all at approximately the same level of proficiency at the outset of the study, an argumentative writing test was administered. The pre-test, immediate and delayed post-tests comprised argumentative writing tasks for all groups. In Study 2, the Metacognitive Strategy Questionnaire (see Appendix A) was administered in the pre-test and post-tests to all experimental groups. Also, six volunteer participants agreed to think aloud as they wrote down essays during the treatment. A detailed description of these instruments and the rationale for their choice will be presented in the following section.

3.4.1. Argumentative essays

The pre-test, immediate/delayed post-tests, and four sessions of writing practice involved the second task of IELTS Academic Writing Module and were sourced from IELTS item banks. The Academic Writing Module assessment is 60 minutes long. It comprises two writing tasks: 150 words for the first task (20 minutes) and 250 words for the second (40 minutes). In Task 1, candidates are asked to describe some visual information (graph, table, chart, diagram), and to present the description in their own words. In Task 2, candidates are presented with a point of view, argument or problem. The rationale for the selection of

this test was its general acceptance as a reliable benchmark for English proficiency that is currently used in more than 7000 institutions in over 135 countries. According to information on the IELTS website (www.ielts.org), it is the world's most widely-used high-stakes English proficiency test. A considerable number of samples of IELTS writing topics are readily obtainable over the internet or in published books.

The tasks were designed to assess their ability to discuss a problem, point of view or argument. In other words, participants were required to present and justify their opinions, to give solutions to a problem, or to compare differing ideas or viewpoints. They were asked to give reasons for their arguments they included in their writing, and to include any relevant examples from their own knowledge or experience. As regards the writing topics, it was important to choose topics which would relate to learners' academic and social interests where possible. The issues raised in these topics were of general interest, and were considered suitable for participants entering higher education. Thus, the target language use situation involved writing academic essays in L2 in an academic domain. The writing topics required participants to have an awareness of formal register and the rhetorical conventions of argumentative genre. These topics are set out in Table 3.2 below.

Table 3.2 Writing topics

Writing Task	Writing Topics
Pre-test	The idea of going overseas for university study is an exciting prospect for many people. While it may offer some advantages, some would rather stay home because of the difficulties a student inevitably encounters living and studying in a different culture. Compare these two views. Which view do you agree with? Why?
Session 1	In some countries, marriages are arranged by parents, but in other cases, people choose their own marriage partners. Discuss both systems and give your own opinion.
Session 2	Some people think that children should begin their formal education at a very early age and should spend most of their time on school studies. Others believe that young children should spend most of their time playing. Compare these two views. Which view do you agree with? Why?
Session 3	Some parents raise their children in strict discipline, while others leave them free to learn lessons of lives on their own. Discuss both approaches and give your own opinion.
Session 4	Successful sports professionals can earn a great deal more money than people in other important professions. Some people think this is fully justified while others think it is unfair. Discuss both these views and give your own opinion.
Immediate Post-test	Some people argue that it is more important to have a single language to be adapted as the official international language. Others think that this will lead to the loss of culture and identity of one's country. Discuss both these views and give your own opinion.
Delayed Post-test	Some people believe that a college or university education should be available to all students. Others believe that higher education should be available only to good students. Discuss these views. Which view do you agree with? Explain why?

Note. Retrieved in June 2010 from:
www.scribd.com/doc/5128744/ielts-writing-samples-task-2
www.ielts-exam.net/index.php?Itemid=28

These seven topics were selected from a large number of sample topics retrieved from the two links listed above. Initially, 14 topics were piloted among five L2 learners who were asked to rank the topics according to their perceived level of difficulty. The seven topics ranked as the easiest to write about were selected as the writing topics for this study. Of these seven topics, those three ranked as more difficult were selected for the pre-test, immediate and delayed post-tests. Similar to Manchón and Roca de Larios's (2007) study

on investigating the temporal nature of planning in L1 and L2 composition, because there was a risk of participants in each group disclosing the prompts to participants in other groups, the same writing topics were administered to participants in all groups and the order of the writing tasks was maintained constant across all four experimental groups.

Since the writing texts in the current study are samples of IELTS Academic writing Task 2, a brief description of the test characteristics, structure, and band scales is considered essential. In the IELTS writing module, depending on whether the candidates opt for either the General or Academic versions of the test, two different pieces of writing are produced. In both types, the writing module is divided into two parts and the candidates are allowed 60 minutes to complete two tasks, of 150 and 250 words. The IELTS General Writing, which is intended for immigration purposes or to undertake non-academic training, comprise two tasks, one always being a letter, the second being an essay based on a given topic which is of general, rather than of specialist, knowledge interest. The IELTS Academic Writing, which is a requirement for obtaining admission to secondary, tertiary institutions of higher education, vocational and training programs, particularly in many English-speaking universities like British, Australian, New Zealand and Canadian universities, has also two significant writing components. In the first task, the candidates are given some visual information which may be presented in the form of one or more related diagrams, graphs, charts, or tables. In the second task, the candidates are presented with an opinion, problem or issue which they must discuss. They may be asked to present the solution to a problem, present and justify an opinion, compare and contrast evidence or opinions, or evaluate and challenge an argument or idea.

All modules of IELTS are scored on a nine-band scale, with each band corresponding to a specified competence in English (see Appendix D). These nine bands include detailed descriptors in the defined rating scale of IELTS to distinguish between different levels. An analytic rating scale with four distinct categories is administered in scoring the writing tasks. These categories focus on salient characteristics of writing and are defined in concrete terms at different nine bands (see Appendix E). This rating scale is intended to increase the inter-rater reliability and to better discriminate between different writers.

To allow for more differentiation within a band, the scores of the writing module have been reported in whole or half-bands since July 2007. Once each scale is scored by the

rater, a combined score of the four sub-scores is then reported. A rounding conversion tends to apply: if the average score of the sub-scores ends in .25, it is rounded up to the next half band, and if it ends in .75, it is rounded up to the next whole band. The IELTS exam website (www.ielts.org) provides more detailed information on the test formats, sample questions and preparation materials.

Procedures

Prospective participants were first asked to write an argumentative essay which was a sample of the IELTS writing task two. In Study 1, I used IELTS detailed descriptors to independently award a band score for each of the four criterion areas which were equally weighted: *task response, coherence and cohesion, lexical resource,* and *grammatical range and accuracy*. This rating was done by me and a certified rater in Study 2. In order to select appropriate participants, a band score minimum of 5 and maximum of 6 was set prior to the study. Due to the higher level writing skills required in argumentative genre of writing, the selection of this range of band scores ensured that the participants have the pre-requisites of linguistics knowledge and thus do not struggle with linguistic encoding of their message.

Based on the results of the proficiency test and after ensuring the proficiency comparability of participants at the outset of the project, 35 participants were selected from among 44 in Study 1 and were randomly divided into four homogenous groups, as shown in Table 3.3. The 70 participants in Study 2 were chosen from among 100 and were randomly divided into five homogenous groups. One control group (n = 15) took only the pre-test, immediate and delayed post-tests, and four experimental groups undertook the treatment sessions depending on the groups they had been allocated to. Study 2 included a control group in order to better compare the impact of the treatment on improved writing performance. Therefore, there were four experimental conditions, with two levels of pre-task planning: pre-task planning (+P) and no pre-task planning (-P), and two levels of metacognitive training: metacognitive instruction (+M) and no metacognitive instruction (-M).

Table 3.3 Group conditions

		Metacognitive Instruction (+M)		**No Metacognitive Instruction (-M)**	
		Group	***n***	**Group**	***n***
Study 1	**Pre-task Planning (+P)**	+M+P	*10*	-M+P	*8*
	No Pre-task Planning (-P)	+M-P	*10*	-M-P	*7*
Study 2	**Pre-task Planning (+P)**	+M+P	*14*	-M+P	*14*
	No Pre-task Planning (-P)	+M-P	*13*	-M-P	*14*
	Control group (*n = 15*)				

Note. M = Metacognitive instruction, P = Pre-task planning

In the standard IELTS test, only 40 minutes is allowed for the second writing task. In this study, however, a time limit of 45 minutes was allowed, based on previous piloting of the task with six participants who were given an unlimited time to complete the task. The range of time taken was 35 to 55 minutes, with an average of 45 minutes. Thus, the time required to complete the task in both studies was set at 45 minutes. In groups +M+P and -M+P, the participants were required to write at least 250 words within 45 minutes, during which they were also given a sheet of paper to write notes for maximum ten minutes before they started the writing task itself. They were told, however, not to write out whole sentences of the essays in their plan, but they were allowed to start writing before ten minutes had elapsed. The choice of ten minutes of planning time followed the studies of Ellis (2003), Ellis and Yuan (2004), Shin (2008), and Ojima (2006) which explored the impact of planning in writing. This period was found to be sufficient to favour the quality of written language production. Similarly, in the case of oral production, Mehnert's (1998) study showed that only when at least a ten-minute planning time was provided was there a measurable effect on all three aspects of fluency, accuracy, and complexity. The main difference between the two groups was that group +M+P was taught planning strategies as the first metacognitive strategies in the training. Group -M+P, however, received no detailed guidance, and was just requested to plan their writing in terms of content, organisation, and language. In groups +M-P and -M-P, as in the other two conditions, participants were given their writing sheets and requested to finish writing within 45 minutes and to produce at least 250 words. These groups, however, were not given ten minutes to plan before beginning the task. I ensured that they began writing immediately.

One major difference between this study and other studies conducted on pre-task planning (e.g. Ellis & Yuan, 2004; Shin, 2008; Zhigang & Xudong, 2008) is that the notes made by students during the planning time were not removed before participants started the writing task. Ellis & Yuan (2004), for instance, maintained that removing the notes ensured that the language elicited by all tasks was produced within the specified time limit. Only in Ong and Zhang's (2010) study, +/– draft availability was manipulated as a variable to explore task complexity in EFL students' argumentative writing. In the current study, however, the time limit for all groups was equal (45 minutes) and the time two of the groups spent on taking notes during the pre-task planning was not an additional time allowance. In other words, pre-task planning took place within the allotted period of 45 minutes. In contrast to Ellis & Yuan's (2004) study in which removing pre-task planned notes in narrative tasks did not seem to hinder writing, I argue that producing more complex tasks such as argumentative essays without access to the previously produced drafts as an aid in planning what to write next can prove to be comparatively more difficult.

The participants in all groups were first asked to write an argumentative essay, as a pre-test. These essays were analysed to calculate the complexity and accuracy of participants' initial task-based written performance, and to assign a holistic score based on IELTS nine-band writing scale. The training program with four sessions of instruction carefully devised for each experimental group was carried out over a period of four weeks. Each group produced a practice argumentative text at the end of each instructional session. One week after the four training sessions had finished, participants were asked to write another argumentative writing as a post-test. In Study 2, participants were required to write another essay as a delayed post-test three weeks after the immediate post-test. The texts they produced were then analysed with regard to complexity and accuracy, and were also given a holistic rating score.

3.4.2. Questionnaires

In Study 2, a Metacognitive Strategy Questionnaire (MSQ) was administered after the writing task in the pre-test, immediate post-test and delayed post-test sessions (the full items of the questionnaire can be seen in Appendix A). This type of questionnaire was piloted with participants in Study 1 to check for appropriate wording, clarity, ease of use,

and consistency with the purpose of the study and the classification of metacognitive strategies. The questionnaire data obtained in Study 2 were used to help increase the validity of study findings. The primary goal of the questionnaire was for the participants in the four experimental groups in Study 2 to report their perceptions of the metacognitive strategies they had used in their argumentative writing.

Development of MSQ

The MSQ was adopted and substantively modified from Purpura's (1999) 80-item Cognitive and Metacognitive Use Questionnaires and Tsai's (2004) Cognitive (20 items) and Metacognitive (14 items) Strategy Use Questionnaires. They both attempted to achieve a single model which could possibly account for both cognitive and metacognitive strategy use. The questionnaire in this study, however, drew mainly on the metacognitive strategies of the questionnaires devised by Purpura (1999) and Tsai (2004). I also consulted other metacognitive strategy taxonomies in the literature and compiled a list of metacognitive strategies for inclusion in the questionnaire and wrote operational definitions for each strategy. Then I developed items to measure each classification of strategies in writing. It is important to note that a number of task-specific questionnaires have been constructed to examine metacognition. For instance, the items in Samuelstuen and Bråten's (2007) questionnaire were tailored in conjunction with an expository reading text about socialisation. Richardson (2004) and Samuelstuen and Bråten (2007) argue that, in questionnaires on general learning strategies, learners are required to access their long-term memory, within which they have abstracted general characterisations of their learning strategies over multiple occurrences. However, as Veenman (2011) has argued, whether or not they have retained a record of the required mental activities accurately in their long-term memory is questionable. Contrary to a number of questionnaires such as Metacognitive Awareness of Reading Strategies Inventory (MARSI) constructed by Mokhtari and Reichard (2002) which is a theory-based general measure of strategy, the context of the items in MSQ questionnaire was precisely defined for the specific writing task the participants just completed as a frame of reference. In MARSI, the items were asking "What do you generally do when…", and thus the items did not refer to a specific or detailed reading task, whereas the type of statements in MSQ questionnaire specifically aimed at assessing the activities applied in the context of academic writing the participants just completed. For instance, in the third category of the questionnaire, the participants were asked to contemplate on the statement "When I was writing, I tried to think about…"

So the MSQ questionnaire was designed to assess the participants' awareness and actual use of the writing strategies pertaining to the writing task they just completed.

The final version of the MSQ included statements which aimed at addressing the participants' activities during actual writing. Participants were instructed to rate their degree of agreement or disagreement with the activities occurring in their writing situation. The participants' rating of the statements was done on a six-point Likert-type scale, with "1" indicating strong disagreement and "6" expressing strong agreement. I decided that the six-point Likert alternatives (i.e. Strongly Disagree, Disagree, Slightly Disagree, Slightly Agree, Agree, Strongly Agree) may be better suited to my research purpose than a scale with a neutral or impartial mid-point (i.e. Strongly Disagree, Disagree, Neither, Agree, Strongly Agree). When deleting the neutral response, the participants were required to decide whether they leaned more towards the (strongly) agree or (strongly) disagree end of the scale for each item. Plus, since the questionnaire items were crafted to suit writing task specificity, I considered a more elaborate 1-6 scale would help the participants to pinpoint their degree of agreement/disagreement with more precision. Hereto, more accurate and discriminating responses were aimed to be collected.

Four types of strategies were represented in the MSQ: planning, considering the audience, monitoring, and evaluating. Planning strategies were ascertained via five items that addressed participants' conscious attention to the preparatory steps prior to undertaking the task. Responses to these questions were used to indicate participants' perceptions about planning. Three items on the questionnaire called for the participants' awareness of a perceived audience while writing. Participants were asked to make evaluative judgments about their sensitivity to the audience on a scale from one to six. The questionnaire contained eight items related to monitoring strategies, and four items related to evaluating strategies.

Procedures

In Study 2, the participants in the four experimental groups were asked to complete the Metacognitive Strategy Questionnaire at three phases of time, i.e. pre-test, immediate post-test, and delayed post-test, to help measure their reported writing strategies which were specific to the domain of argumentative writing. The questionnaire administered at the pre-test session attempted to ascertain participants' initial degree of awareness in the four

metacognitive domains. This established a baseline from which to evaluate their perceived progress after the treatment, which was measured via re-administration of the same questionnaire at the post-test sessions. This post-test (immediate and delayed) questionnaire aimed to measure any perceived changes from before the start of the treatment to the completion of treatment. It took them around ten minutes to complete the questionnaire.

3.4.3. Think-aloud protocols

Concurrent verbal protocols, also referred to as think-alouds, provide concurrent real-time data as opposed to post hoc verbalisation. They have been increasingly used by qualitative researchers interested in obtaining a rich and valid source of data for process inquiry tasks such as writing (Hayes & Flower, 1983; Janssen, Van Waes, & Van den Bergh, 1996; Manchón & Roca de Larios, 2007; Manchón, Roca de Larios, & Murphy, 2000; Ransdell, 1995; Roca de Larios, Murphy, & Manchón, 1999; Sasaki, 2004). It is, however, generally acknowledged that there are some inherent problems in the use of think alouds. For instance, it is argued that it is difficult for some participants to think aloud while writing, especially when the writer is verbalising in L2 (Ericsson & Simon, 1993). Furthermore, many L2 learners may be accustomed to thinking in their L1 while writing. Another problem is that think-aloud conditions may affect the quality and content of the participants' cognitive activities while writing. For example, some participants may be more focused on the verbalisation, and this may disrupt the sequence of their writing (Van Someren, Barnard, & Sandberg, 1994). Due to the additional time that was required to produce the overt verbalisation of thoughts, the think-aloud participants in the present study took longer to complete the tasks. There were times where participants were merely verbalising one thought, but switched to others as they emerged. This additional cognitive activity could at times change the sequence of meditative thoughts.

All in all, and in spite of the limitations reported above, think-aloud techniques have been praised by some scholars. The rigorous method of protocol analysis allows researchers to elicit verbally reported sequences of thought. These techniques have proved "extremely productive", Hyland (2010) argues, "in revealing the strategies writers use when composing, particularly what students do when planning and revising texts" (p. 197). Similarly, the purpose of selecting this instrument in the current project was to help

consolidate the processes that control the generation of new thoughts in writing and the choice of alternative procedures and solutions to novel tasks in each session. In order to strengthen and corroborate data collected from the writing essays and questionnaires and to bring greater plausibility to the interpretation of results, think-aloud protocols were used in Study 2. The purpose of collecting these data was not, however, to generalise the in-depth analysis of these protocols to a broader population, but to characterise the similarity and diversity of individual strategy use within the context of text production.

Procedures

In Study 2, two male and female participants in group +M+P, one female in group +M-P, one male in group -M+P and two males in group -M-P were recruited. The recruitment of volunteers for this activity took place at the start of the pre-test session, when the nature and purpose of the think-aloud activity was explained. In order to demonstrate composing aloud with a topic and prepare the volunteer participants for the activity, the recorded tape of one participant in Study 1 who was asked to think aloud as a pilot was played to them. Before the first session, participants were briefly instructed to verbalise their thinking while writing. In order to establish their familiarisation with the think-aloud technique, they were also given the opportunity to practise with a mock composition for about 15-20 minutes at a separate session. The choice of briefing time was based on the time allowed in Manchón et al.'s (2000) study.

Participants were encouraged to verbalise what came into their minds as they composed. Because there was a concern about the possibility of intrusive effects, participants were instructed to write in their usual way in every session of the workshop, and in the same way as they had done in the pre-test, but to verbalise their thoughts while carrying out the writing task. They were told that this was so that the researcher could further investigate their strengths and weaknesses in writing and analyse their writing process from information collected about their decisions, strategies and perceptions. Think-aloud protocols, which were the L2 verbalisations of these participants' thought processes, were recorded on audiotape. I then fully transcribed the audio-taped protocols. Two of the participants asked for electronic copies of their recordings and final transcripts.

Summary

This section accounted for the choice of instruments, corroborated by the written texts as the main instrument of data collection and the questionnaires and think-aloud protocols as complementary instruments which helped the researcher to gain access to rich information on participants' thought processes. It also explained procedures used to conduct the study and process the resulting data. A summary of the data collection procedures for Study 1 is diagrammatically demonstrated in Figure 3.3.

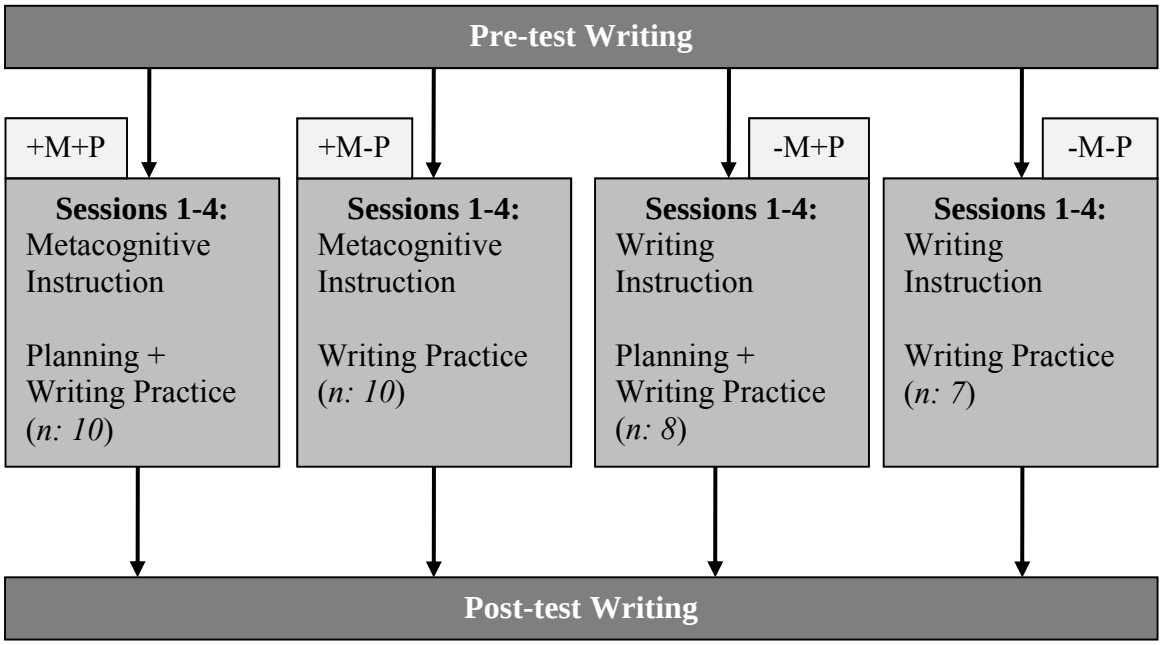


Figure 3.3 Data collection procedure in Study 1

Overall, Study 2 followed the same procedure as in Study 1 with the following additions and adaptations:

1. A fifth group was added as a control group. This group completed all pre- and post-tests, but did not participate in treatment sessions.
2. 70 participants took part in the study.
3. A delayed post-test was added to evaluate any long-term and transferable effects of the instructions, i.e. the performance was checked over an extended period in addition to an immediate post-test.
4. Six participants took part in the think-aloud protocols.
5. The Metacognitive Strategy Questionnaire was administered to all experimental groups in the pre-test and post-tests to ascertain whether the treatment had resulted

in any changes in participants' strategy awareness and use, as compared with the initial administration of the questionnaire in the pre-test.

A summary of the data collection procedure for Study 2 is presented in Figure 3.4.

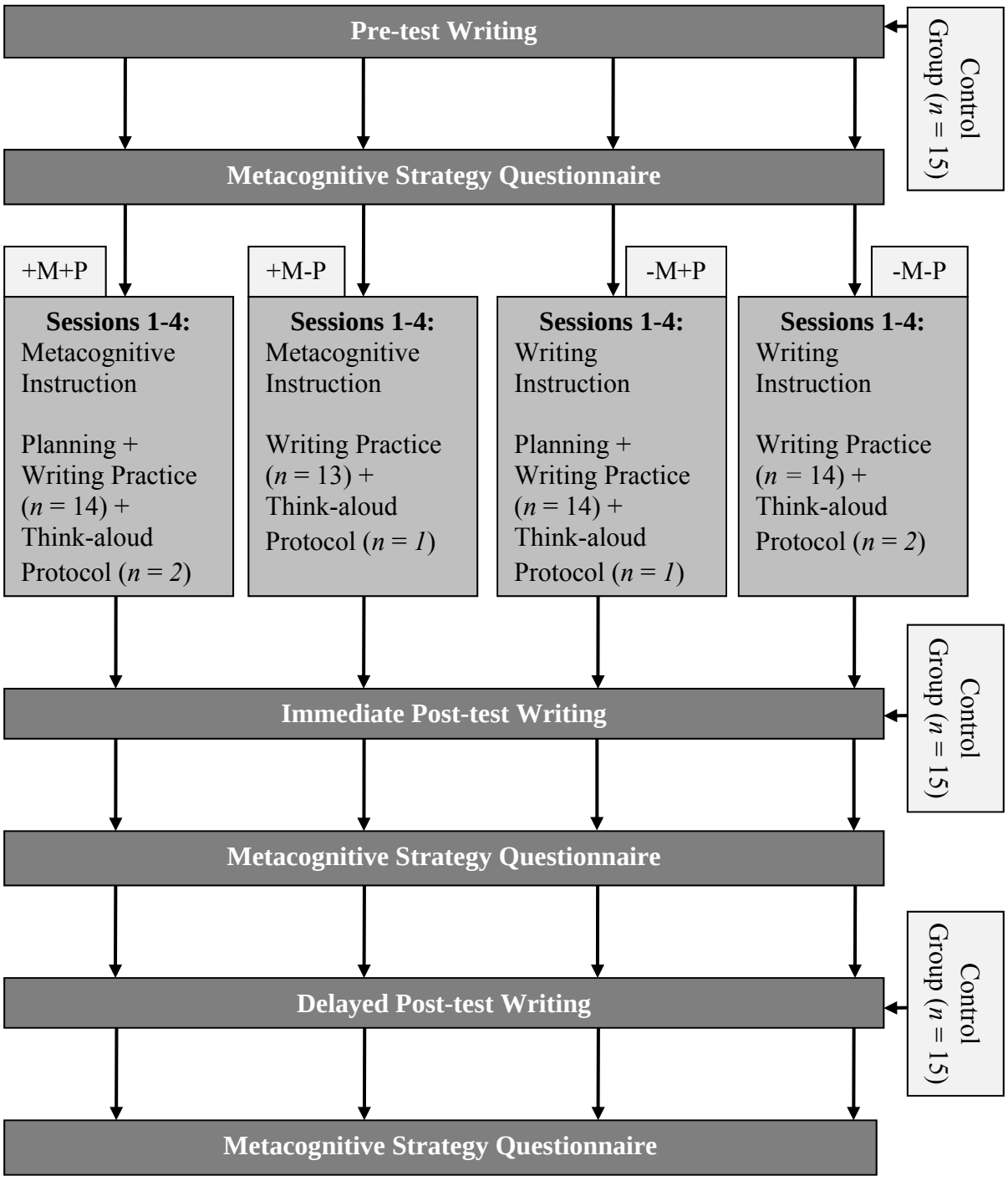


Figure 3.4 Data collection procedure in Study 2

3.5. Unit of instruction

Following the pre-test session, in which participants' writing ability was evaluated, four instructional sessions took place. Each session lasted for around two hours: one hour of instruction, five-minute break time, five minutes to distribute the writing sheets, and forty-five-minute writing task.

Two of the experimental groups (+M+P and +M-P) received metacognitive training. This included four sessions of writing instruction to teach these two groups elements of academic essay writing and metacognitive strategies. The other two experimental groups (-M+P and -M-P) were taught the same elements of academic essay writing and were familiarised with other rhetorical modes of discourse in essay writing, but received no metacognitive training. Each experimental group produced an argumentative text after each session of instruction.

The instruction offered to groups +M+P and +M-P was the same, except that Group +M+P was given the opportunity to plan for ten minutes prior to writing, while Group +M-P was asked to start the writing task without any pre-task planning. The instruction for groups -M+P and -M-P was also identical, except that group -M+P had the chance to plan for ten minutes prior to writing while Group -M-P started the writing task with no pre-task planning.

As shown in Table 3.4, all four groups were taught the same elements of essay writing. However, the instructional content of each session for groups +M+P and +M-P was designed in such a way as to specifically address writing metacognitive strategies while different rhetorical modes of discourse in essay writing were the main focus of attention in the instruction given to groups -M+P and -M-P.

Table 3.4 Instructional treatments

Sessions	All Groups	Groups +M+P +M-P	Groups -M+P -M-P
	Elements of Essay Writing	**Writing Metacognitive Strategies**	**Rhetorical Modes of Discourse in Essay Writing**
1	Introduction Paragraph	Planning	Narrative Essay
2	Body Paragraph (Support)	Considering the audience	Descriptive Essay
3	Body Paragraph (Refutation)	Monitoring	Expository Essay
4	Conclusion Paragraph	Evaluating	Argumentative Essay

Note. P: Planning, M: Metacognitive instruction

Based on the specific content to be covered in each session, four detailed lesson plans were designed (see Appendix B for a sample lesson plan for the first session of the metacognitive training). In order to enhance the effectiveness of the sessions, PowerPoint slides were created that included texts and visual components. Using the same graphic consistently for the introduction, maintenance and practice of some ideas served as visual cues that helped participants connect the concepts presented (see Appendix C for sample PowerPoint slides in the first session for the metacognitive training). What follows is a brief summary of the content of each of the four sessions, both the common content of instruction for all groups and the content particular to writing instruction with or without metacognitive training.

3.5.1. Common instructional intervention

The content of almost half of each session was the same for each group of participants. The purpose was to help participants to develop each individual paragraph of an essay effectively.

Session 1: Introduction Paragraph

Within the first session, the participants were taught how to develop an introduction paragraph. They were initially introduced to what Harris, Santangelo, and Graham (2010) describe as the mnemonic TREE: *Tell* what you believe (e.g. Topic sentence), provide two or more *Reasons* (e.g. Why do I believe this? Will my reader believe this?), *End* it, (i.e.

Wrap it up right) and *Examine* (i.e. Do I have all my parts?). After an overview of the purpose of paragraphing had been introduced, the essentials and components of an introductory paragraph were discussed. Finally they were prepared as best for the first writing practice of the session.

Session 2: Body Paragraph (Support)

Participants worked on the structure of the body paragraph: (a) Point (Topic sentence: proves thesis), (b) Claims, (c) Supporting evidence. Participants were advised to allocate a paragraph to each main point, supported by evidence, so that each paragraph pertained to one key point only with everything in the paragraph supporting that one main point. The participants and I discussed how writers most often would begin with the weakest point and conclude with stronger arguments. They agreed that if all the points were equally strong, they would build a relationship with their audience by discussing the most familiar, less controversial points first. It was emphasised that when making claims, there are different ways to provide supporting evidence, e.g. facts, numbers, statistics, reliable research, relevant data, illustrative examples, and statements by authorities, all of which would help the arguments seem convincing. Exemplification was introduced as another way to provide evidence. To do so, participants were encouraged to include short narratives and descriptions based on personal experience or the experiences of others. To give a better clue of body paragraphs with claim and evidence, sample paragraphs containing extended and hypothetical examples were distributed.

Session 3: Body Paragraph (Refutation)

In order to familiarise participants with the concept of refutation, they were asked to draw a table and label one column *NO* and the other column *YES*. They were then asked to mark an X representing where they would stand on a controversial statement read to them: "Students should wear uniforms in school". That is to say, if they agreed, they would draw an X on the *YES* end of the table; if they disagreed, they would draw an X on the *NO* end; if they were undecided, they would draw an X somewhere in the middle. Next, they were asked to write down three reasons why they marked X at the *YES* or *NO* end. When they finished writing their reasons, a volunteer from each viewpoint was asked to stand up and refute the opposing arguments. In turn, a volunteer student offered a point in support of his/her position, and a student with an opposing view refuted and offered a supporting reason for his/her viewpoint. The researcher then summarised the task by giving remarks

on the whole task and focusing on particular words, phrases, or sentences the volunteers used to attempt to win the battle (e.g. I totally disagree with you; I should argue that your point has nothing to do with this issue, etc.). It was pointed out that although the word *argumentative* could be confused with having *arguments*, in academic argumentative writing the writer needed to address both supporting and opposing arguments and concede merits in other points of views.

To develop and practice skill in this kind of writing, participants considered a number of viewpoints with regard to the topic of the last session writing task. The researcher explained that writers needed to show that they understood and genuinely respected their readers' positions even if they thought the position was ultimately wrong, because discussing an opposing viewpoint and acknowledging the difficulty of making absolute value judgements in extreme positions and developing a compromise position would add strength to an essay. In fact, by indicating awareness of an opponent's point of view, the writer could then refute these arguments, showing how the writer's own ideas were more valid. Even if dissatisfied with someone else's explanation of a phenomenon, analysis of an event, or solution to a problem, they would have the opportunity to write a paragraph called refutation, in which the writer would acknowledge and disagree intelligently with the opposition. This acknowledgement, however, did not imply agreement. Instead, refutation would mean disagreeing intelligently with the opposition.

Session 4: Conclusion Paragraph

In the fourth session, different strategies of concluding an essay were practiced. These included: restating the major thesis, summarising key points, reinforcing the weakness of one's opposition, underscoring the logic of one's presentation, re-emphasising why that debate was important, suggesting a course of action, or challenging readers to apply the essay's argument to their own life.

3.5.2. Writing plus metacognitive instruction

An important issue for researchers on metacognition is whether metacognition is general by nature and thus can be instructed concurrently in different learning situations, or task and domain specific and thus needing to be taught for each task or domain separately (Borkowski, 1992; Veenman et al., 2006). My research pertains to "task-specific

metacognition" which was taught to address academic writing tasks. Having carefully designed the lesson plans and the PowerPoint slides, I decided on my choice of words, phrases, metaphors and interaction sequences to invoke a thoughtful classroom. Words such as *think*, *imagine*, *reflect*, *hypothesise*, *predict*, *guess*, *expect*, *explain*, *support*, *classify*, *clarify*, or *justify* were frequently used. This meticulous attention to the language used to talk about thinking was advocated by Kolencik and Hillwig (2011) who argue that teacher's language of thinking can help students to regulate their thoughts about learning. When formulating questions, in particular, they urge teachers to use terms and phrases which require students to justify and support their views rather than providing a quick and flippant response (p. 15). Similarly, in encouraging educators to become more mindful of their language of instruction, Costa (2008) illustrated how using specific cognitive terminology rather than vague abstract terms could help students to internalise those terms and even make them a part of their own vocabulary. Prior to teaching each strategy, I raised questions, created dilemmas or posed problems to encourage participants to get involved in the process of problem solving. This followed Costa's suggestions of effective starting points for developing and establishing a thoughtful classroom climate.

In the following section, the instructional treatments within each session of metacognitive training (with or without planning) are presented in detail. For each regulatory strategy, information about how to use the strategy, the conditions under which the strategy would be most useful, and the rationale for why to use them while performing the task was provided. The purpose was to promote explicit declarative, procedural, and conditional knowledge about each strategy. These three sets of knowledge are the subcomponents of knowledge of cognition. Strategies like self-questioning were explicitly instructed, modelled and practiced. Table 3.5 shows some examples of self-questions during planning, monitoring and evaluating which were adopted from Hartman (2001, p. 58).

Table 3.5 Sample self-questions (Hartman, 2001)

Strategy	Self-questions
Planning	What is the purpose of this essay? What should be in the introductory paragraph? How should I put these ideas in the best order?
Monitoring	Am I elaborating on all of my main points? How clearly am I expressing my ideas? Am I making any spelling or grammar mistakes?
Evaluating	What was best about my paper? Why did I make those spelling and grammar error? How can I best prevent those mistakes next time?

Session 1: Planning

The aim of the first session was to ensure that participants acquired planning strategies to allow them to successfully manage the writing task. Through whole-class discussions, the reasons why writers would need to plan before starting to write were elaborated. Some possible reasons were: to put together the related and relevant ideas on the topic and develop an organised piece, to remind one of important points that should be covered in the essay, to highlight the final structure of the essay. Participants were then asked what they could plan before starting to write. Different techniques for planning for time, content and language features were discussed and practiced separately.

The participants agreed that sometimes they were required to complete timed writing tasks. Thus managing time and allocating it adequately to each part of the essay was regarded as an important skill in writing an essay. Depending on the whole time for the essay, the topic and length of the essay and the number of arguments to be included in the body paragraph, writers need to allocate an appropriate amount of time for each paragraph. As with planning for the content of written texts, a variety of pre-writing strategies were introduced. One of those techniques was clustering, or also called mind mapping, which was suggested as particularly helpful for those who were visually oriented. With regard to planning for language features, brainstorming was introduced as a possible strategy to generate a list of words and phrases that might be useful in the writing phase. It was emphasised that when brainstorming, it was not necessary to keep writing continuously, but just to jot down ideas that seemed related to the topic and to compile with a list of

words and phrases. To make prewriting activities maximally effective, participants worked on examples of generating ideas through brainstorming and clustering.

Session 2: Considering the audience

During the second session, participants were encouraged to demonstrate sensitivity to the needs and perspectives of their audience. The participants agreed that when they were in the process of writing an essay, it was easy to neglect consideration of the reader of the text. Thus, it was considered important that before they even began the process of writing, they would take some time to reflect on potential readers and the information they needed from them. The possible benefits of keeping the audience in mind were discussed in group work. Possible advantages were making good decisions about what to include, in what order to organise ideas, how best to support argument, how to control the writer's vocabulary, sentence structure, the number of details and the kinds of details.

It was pointed out that by treating their readers as an intelligent but uninformed audience, they would end up addressing them more effectively. In order to further elaborate the concept of "Who is your intended audience?" some possible audiences were introduced: (1) familiar, known audiences: self, friends, peers, family, teachers, (2) extended, known audiences: community, student body, local media, (3) extended, unknown audiences: wider range of media and other publications. It was explained that in academic writing, it would be helpful to think about a specific person toward whom the writing would be addressed. A technique called "character prompts" was introduced to help create a potential audience. Participants were first given topics and asked to imagine a person as an audience with a name, an age, a profession, and a physical appearance. Then some questions to ask in order to create such an audience were elicited. These included: "What is this person's current attitude toward this topic?", "How much does this person know about the topic?", "Describe this person's value system.", "How does this person's value system influence his or her attitude toward the topic?", "Which aspect of the topic does this character find most disturbing?"

Session 3: Monitoring

During the third session, participants were encouraged to monitor their writing performance and behaviour. They were told that when writing, it was essential that one would keep on thinking, all throughout the writing time, of how the essay had been

developed so far and how it would finish smoothly. More importantly, what needed to be attended to more was far more than correcting spelling, grammar, word choice, and punctuation. It involved dealing with the overall content and meaning of writing by adding, deleting, or reorganising larger chunks of discourse. Having given examples, I explained that some revisions would affect the meaning of the text and some would not. Furthermore, some corrections could be at surface (sentence) level and some could be deep (discourse) level corrections in which the writer would check the logicality, relevance of single ideas to the global argumentation pattern, and appropriateness of content to a given reader.

Session 4: Evaluating

In the final session, the strategy of evaluation was introduced and reinforced. The participants were primarily given a sample paragraph which was an uncorrected excerpt from the first draft of a student's paper. They were given some time to read the paragraph and correct any mistakes they could come across. I then listened to the corrections and wrote them on the board. Most corrections were of sentence-level types, e.g. grammatical, spelling, word choice. Accordingly, I explained that in addition to the surface-level corrections, one should not ignore the more fundamental problems of content and depth. This time, the participants were encouraged to go through the paragraph once again, thinking of corrections in terms of content and focus. Later, they were asked to rewrite the paragraph in their own words, having kept the focus of the topic. Thus, I first provided an opportunity for them to better monitor a piece of written work as well as their thinking. Then, I assisted them in developing an ability to self-monitor, and become an independent writer for a new piece of writing.

In summary, the five main components of the explicit metacognitive strategy training in the current study consisted of the following:

1. Selecting the metacognitive strategies employed by successful expert writers (generating ideas via brainstorming, considering the audience and the writing purpose, organising ideas, monitoring the flow of writing, evaluating the appropriateness of content and organisation)
2. Presenting the strategy of planning as explicitly as possible, and training the participants to use some related techniques like idea graphic organisers during the pre-task planning to help the participants to minimise the information-processing demands on their mental resources.

3. Scaffolding the participants during different stages by a number of guiding questions concerning the purpose of writing (e.g. what is the purpose of this writing? How should you present your ideas?), and contemplation of the intended audience (e.g. who is your reader? What does the reader know about your topic? What aspect of your argument may the reader be interested to know about?)
4. Describing, modelling and providing examples of how to potentially monitor and evaluate not only the linguistic aspects, but also the global content and organisation of the written products.
5. Providing ample writing practice opportunities for the participants to apply and integrate the trained strategies to new tasks by a provision of contextualised strategy practice.

3.5.3. Writing minus metacognitive instruction

For groups -M+P and -M-P, different rhetorical modes of discourse in essay writing were introduced within the four sessions: narrative, descriptive, expository, and argumentative. In order to gain familiarity with and an understanding of diverse genre conventions, good models of each type of writing genre were provided in each session of non-metacognitive training (with or without pre-task planning).

Session 1: Narrative essay

The participants were given an example of a narrative prompt. I explained that a narrative essay would tell a story, with characters, setting, and action. I also explained that the characters, the setting, and the problem of narratives were usually introduced in the beginning, the problem reached its high point in the middle, and the ending resolved the problem. I then clarified that the purpose of this type of writing was to recount a personal or fictional experience or to tell a story based on a real or imagined event. In well-written narration, a writer would use insight, creativity, drama, suspense, humour, or fantasy to create a central theme or impression. Thus the details all would work together to develop an identifiable story line that would be easy to follow and paraphrase. Finally the conventions of narrative essays were discussed in class.

Session 2: Descriptive essay

I explained that descriptive essays would be used to create a vivid image of a person, place, or thing. They would draw on all senses, not merely the visual, with the purpose of enabling the reader to share the writers sensory experience of the subject. Thus, descriptive writing portrayed people, places, things, moments and theories with enough vivid details to help the reader to create a mental picture of what was being written about. I brought participants' attention to the elements to consider as they wrote descriptive essays. They were each asked to think of an instance to describe and were encouraged to discuss by multiple questions. I then gave two samples of descriptive essays which were developed either as a description that appealed to the senses or a spatial-order description. The participants worked in groups to distinguish the differences and finally the conventions necessary in each type were discussed.

Session 3: Expository essay

Giving an example of an expository prompt, I explained that such essays could take a variety of forms. It may tell how to make or do something, report on an experience, or explore an idea. Expository writing would convey information to the reader in such a way as to bring about understanding, whether it be of a process or procedure, or of the writer's ideas about a concept. I clarified that the purpose of this type of writing was to inform, clarify, explain, define, or instruct by giving information, explaining why or how, clarifying a process, or defining a concept. Well-written exposition would have a clear, central presentation of ideas, examples or definitions that would enhance the focus developed through a carefully crafted reader's understanding. These facts, examples, and definitions were objective and not dependent on emotion although the writing may be lively, engaging, and reflective of the writer's underlying commitment to the topic. I also brought participants' attention to the elements to consider as they wrote expository essays. They were required to think of a process to explain and were facilitated by the questions I asked. The participants worked in pairs and took notes to produce an expository essay.

Session 4: Argumentative essay

I explained that an argumentative essay would state an opinion and support it convincingly. It would consider the nature of the audience and marshal evidence accordingly. Thus, it would be neither completely objective nor wholly emotional. Instead, it would use the controlled feelings of the writing to persuade the audience. I added that argumentative

writing would argue a point of view when there was an issue over which people could not agree. I then explained that methods of developing an argumentative essay could use a number of text types: definition, description, process analysis, comparison and contrast, and cause and effect. I gave an example of an argumentative prompt which required a teamwork activity of discussing how to write to convince the school principal to accept the student's point of view on the effects of watching television on their grades. Several major functions of argumentative writing (e.g. writing to clarify a stance, to refute opposing viewpoints, to persuade, to clarify issues) and the positions that the participants could take to approach the sample prompt were then discussed.

3.6. Data Analysis

While the sources of data were divided up according to qualitative and quantitative methods, my writing research combined both methods to gain a more comprehensive picture of the complex nature of writing. I triangulated participants' written performance with their verbal reports of thoughts while composing, and self-report questionnaires about their actions, behaviours and strategies during writing. Triangulation also occurred in terms of data analysis and in the organisation of the final report of the results of this study.

3.6.1. Written texts

The rating scales employed in the pre-, immediate post-test, and delayed post-test writing tasks were of two types: (1) IELTS holistic rating of writing (see Appendices D & E) in which writing scores are reported on a 0-9 band scale, with .5 band increment in four areas; Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. The final band for this task is effectively an average of the four marks awarded in these four areas. (2) Measures of specific linguistic and discourse variables, the two which have been figured in SLA task-based research as accuracy and complexity measures.

Accuracy measures (i.e. error-free clauses and error-free verb forms) were adopted from Ellis and Yuan (2004). To code error-free clauses, the data were divided into clauses, and lexical, morphological, and syntactic errors were marked and the percentage of correct clauses was used as a score for analysis. Because the error-free clauses score is a holistic measure of accuracy, a specific measure, i.e. error-free verb forms, was also selected. The

resulting score was the percentage of verbs that were used without any errors in tense, aspect, modality, or agreement.

Complexity was measured through syntactic complexity, syntactic variety, and lexical complexity. The first two measures of syntactic complexity and variety were adopted from Ellis and Yuan (2004). The syntactic complexity measure was mean number of clauses per T-unit. According to Hunt (1965), a T-unit is defined as "one main clause with all subordinate clauses attached to it" (p. 20). Following Young (1995), single clauses, main clauses with their subordinate clauses, two or more phrases in apposition, and fragments of clauses produced by ellipsis were all considered T-units. Syntactic variety was measured based on the total number of different verb forms used, which were differentiated based on tense, aspect, modality, and voice. Finally, lexical complexity was measured using the Giraud index of lexical richness. This index is a modified type-token ratio which is calculated based on the square root of the number of tokens. A great number of recent task-based studies (e.g. Adams, Amani, Newton, & Alwi, in press; Daller, van Hout, & Treffers-Daller, 2003; Gilabert, Barón, & Levkina, 2011; Shiau & Adams, 2011) have chosen the Guiraud Index over the Mean Segmental Type-Token Ratio (MSTTR), mainly because the Guiraud Index corrects for the effect of text length (Vermeer, 2000).

The analysis of fluency was excluded from the analytic measures of the written products. In this study, a time limit of 45 minutes was allocated for the completion of written tasks, therefore the time on task was tightly controlled. Besides, based on Skehan's (1998) Cognitive Approach, there are two distinct language systems: an exemplar-based system, containing linguistic knowledge related to ready-made formulaic chunks of language and discrete lexical items, and a rule-based system, including abstract representations of the underlying patterns of language. Skehan argues that the exemplar-based system is ideally suited for fluent language performance, whereas the rule-based system is best suited for complex or accurate language performance. Given the complex nature of argumentative tasks, I chose to investigate the rule-based system which requires more cognitive processing and whether there would appear a trade-off between complexity and accuracy when increasing task complexity.

The IELTS website (www.ielts.org) and the experimental studies carried out to investigate the reliability of ratings (e.g. Shaw, 2004) show coefficients of .81–.89 for the Writing

Module. Based on test data from 2009, the composite reliability estimate for the academic module was .95 with a SEM of .24, which offers a useful measure for overall test reliability and which was considered as an adequate reason for appointing a single rater rather than two raters for the assessment of participants' writing in Study 1. In other words, I ensured that the results of this particular rating achieved high reliability and thus a second rater was not considered necessary. In Study 2, however, the written texts were rated by the researcher and a qualified and experienced English language specialist who was a certified IELTS examiner. IELTS examiners are required to undergo intensive face to face training and standardisation and to re-certificate every two years and their performance is closely monitored to ensure that they can apply the descriptors in a valid and reliable manner so that global standards are maintained.

The raw scores of participants' written texts were fed into the computer software PASW version 20 for further data analysis, throughout which, the criterion of 95% confidence, or a .05 probability, formed the basis of significance testing. In order to test for normality, I looked for normality both visually in P-P plots and at values that quantify aspects of a distribution, i.e. skewness and kurtosis. The Kolmogorov–Smirnov test was used to double check if a distribution of scores significantly differed from a normal distribution. If the data were sampled from a Gaussian distribution, i.e. normally distributed data, the parametric test of one-way independent ANOVA was used. Where the differences proved to be significant, Tukey post hoc tests, consisting of pairwise comparisons of all different combinations of the treatment groups, were conducted. However, non-parametric or distribution-free test of Kruskal-Wallis one-way analysis of variance was used when the assumption underlying normal sample distribution was not met. As for a non-parametric version of a post hoc test, the Mann-Whitney U test was used to make pairwise comparisons between groups, ending up with six comparisons between each pair of treatment groups. To control for the Type I error rate (in which the results are declared to be significant when in fact they are not), a Bonferroni correction was applied, with adjusted alpha levels of .0083 per test (i.e. .05 divided by the number of tests: 6), instead of using .05 as the critical value for significance for each test. Whenever a test statistic was significant, the size of the effect being tested, that is to say *r* as an effect size measure, was reported in an objective and standardised way. What constituted a large or small effect was based on Cohen's conventional criteria (1992):

d =.10 (small effect): In this case the effect explains 1% of the total variance.

d =.30 (medium effect): The effect accounts for 9% of the total variance.
d =.50 (large effect): The effect accounts for 25% of the variance.

3.6.2. Questionnaires

As regards the Metacognitive Strategy Questionnaire, first the reliability (Cronbach's alpha) was established for the full questionnaire. The overall internal consistencies for the subscales were also calculated. Then the descriptives per each four scale and for each four group were computed. Finally additional data analysis including ANOVA and repeated measures analysis were carried out to show between-group and within-group comparison.

A validation process involving both qualitative and quantitative methods was undertaken using test taker as well as non-test taker participants. First, the preliminary questionnaire was presented to my supervisors and a PhD colleague at my department in the University of Auckland. They provided expert judgment regarding the initial classification of metacognitive strategies and the wording of the questionnaire items. Their recommendations resulted in revising, addition and deletion of items. Then the questionnaire was piloted on the participants in Study 1. The intelligibility of the questionnaire items was checked by having two students in Study 1 complete the questionnaire while thinking aloud. The results of the reliability analysis led to further revision of the questionnaire items to take place. These validations resulted in a revised questionnaire which contained a total of twenty Likert-scale items to gather information related to the four areas of metacognitive strategy. The results of the internal consistency reliability of the revised version of the questionnaire are reported in the following.

After the MSQ was constructed in the beginning stages of the instruments development, it was piloted on the participants in Study 1 and an item analysis was conducted. The reliability of the questionnaire in Study 1 was evaluated to be $\alpha = .71$ as a whole. The results of the reliability analysis showed that a number of items yielded low Corrected Item-Total correlations and thus were dropped or reworded to better reflect the strategies. Having made the required alterations to the actual wording of the items, it was hoped that the psychometric quality of the questionnaire would be more fine-tuned in Study 2. Given that only the questionnaire data in Study 2 were analysed to answer the fourth research

question, a detailed report of the reliability analysis of these data is presented in what follows.

According to Dörnyei (2003), the attribute of internal consistency in examining the reliability of questionnaire data refers to "the homogeneity of the items making up the various multi-item scales within the questionnaire. (p. 110). The construct of the Likert-scale questionnaire consisted of four subscales, namely Planning, Considering the Audience, Monitoring, and Evaluating. The reliability of the instrument was evaluated using Cronbach alpha coefficient, and was found to have an Alpha of .76 as a whole. This was considered satisfactory for the four scales with items ranging from 3 to 8. A coefficient of 0.6 is considered the minimum requirement for reliability in second language acquisition research. Any reliability coefficients lower than this minimum would sound "warning bells" (Dörnyei, 2003, p. 112; Dörnyei, 2007, p. 207). The analyses of the individual subscales show that all subscales produced alpha coefficients of at least 0.6. The results of the analyses are presented in four tables (Tables 4.2 to 4.6). These tables show the alpha coefficient for each subscale, the correlation between each item and the total score for each subscale, and the effect of deleting each item on the alpha value.

Table 3.6 shows the Cronbach alpha for the Planning subscale to be .60 which exactly corresponds with the $\alpha = .6$ requirement for reliability. The correlation coefficients in the Corrected Item-Total Correlation for items 1, 2 and 5 show that the items correlate reasonably well with the subscale itself, since item correlations with the overall score should not be less than .3 (Field, 2005, p. 672). Items 3 and 4, however, did not correlate very well with the subscale, with coefficients of .25 for both items. This could mean that the two items were not good measures of planning for these participants. The values in the column labelled Cronbach Alpha if Item Deleted, however, indicate that none of the items would increase the reliability if they were deleted because all values in this column are less than the overall reliability of .60.

Table 3.6 Reliability analysis: Planning subscale in MSQ

Item	Description	Corrected Item-Total Correlation	Cronbach Alpha if Item Deleted
	Before I started writing,		
1	I planned for the content of my essay.	.49	.46
2	I planned how I was going to structure different parts of the essay.	.46	.47
3	I planned what language features I was going to use in my essay.	.25	.59
4	I thought about how much time I should spend on the essay.	.25	.59
5	I thought about the length of my essay.	.31	.55
	Cronbach Alpha for the 5 items = .60		

It should be noted that the number of items making up Considering the Audience subscale was the fewest among all scales (n = 3). Dörnyei (2003, p. 111) asserts that short scales tend to display more evidence of homogeneity then long ones. Similar to the earlier subscale, however, this scale has a Cronbach Alpha that is a little above .6 (α = .61). Table 3.7 demonstrates that all the items have a correlation coefficient that is above the .3 benchmark, except for item 7 which is below the minimum value required for reliability (α = .28), possibly because it is relatively a general statement. Cronbach's Alpha if Item is Deleted indicated that this item would increase the reliability if it were deleted because the value is more than the overall reliability of .61 (α = .68).

Table 3.7 Reliability analysis: Considering the audience subscale in MSQ

Item	Description	Corrected Item-Total Correlation	Cronbach Alpha if Item Deleted
	When I wrote, I tried to		
6	think about what questions my reader might ask about the topic.	.43	.52
7	think about my reader's views and opinions on the topic.	.28	.68
8	choose the level of formality of my essay to suit the reader.	.61	.22
	Cronbach Alpha for the 3 items = .61		

In Table 3.8, it can be seen that the Cronbach Alpha for the eight items in the Monitoring subscale is .69, which is the highest coefficient obtained amongst the four subscales. Judging from the Corrected Item-Total Correlation, items 9, 10, 11, 13, 14 and 15 had a reasonably good correlation with the subscale but items 12 and 16 did not correlate too well with the other items. None of the items substantially affected reliability if they were deleted, except for item 12. Deleting this item would increase the α from .69 to .70. Nevertheless this increase is negligible and both values reflect a good degree of reliability. This means that almost all the items correlated reasonably well with the subscale.

Table 3.8 Reliability analysis: Monitoring subscale in MSQ

Item	Description	Corrected Item-Total Correlation	Cronbach Alpha if Item Deleted
	When I was writing, I tried to think about		
9	how much time I had remaining.	.39	.66
10	whether I was spelling some words correctly.	.31	.67
11	whether I was using appropriate vocabulary.	.45	.65
12	whether I was using the correct grammar (e.g. tenses, prepositions,…)	.12	.70
13	how many arguments I should have in the essay.	.54	.62
14	whether the arguments followed the instruction of the essay.	.61	.61
15	what parts my essay should have.	.40	.65
16	how to connect different parts of my essay.	.26	.68
	Cronbach Alpha for the 8 items = .69		

As shown in Table 3.9, the alpha for the Evaluating subscale is .62 which ranks the second highest amongst all four subscales. All items showed a strong correlation with the subscale. The Corrected Item-Total Correlation values for all the items were higher than the .3 benchmark and none of the items would increase the reliability if they were deleted. This indicates that all items positively contributed to the overall reliability.

Table 3.9 Reliability analysis: Evaluating subscale in MSQ

Item	Description	Corrected Item-Total Correlation	Cronbach Alpha if Item Deleted
	After writing, I reread my essay and		
17	made sure the language of my essay was clear.	.31	.61
18	made sure the organisation was easy to follow.	.36	.58
19	made sure I had covered the content fully.	.43	.52
20	made sure all the paragraphs were relevant to the topic.	.50	.47
	Cronbach Alpha for the 4 items = .62		

In sum, all subscales achieved an alpha value of .6 or higher, with Monitoring subscale achieving the highest value of all. Thus, it is evident that almost all subscales have contributed satisfactorily to the overall reliability of the questionnaire. Yet, it is essential to demonstrate how individual items correlate with the overall set of items in the entire questionnaire. To this end, Table 3.10 shows the correlation coefficients of each individual item to the overall score of the questionnaire. As mentioned earlier, there was a satisfactory reliability value for the whole questionnaire (α = .76).

Table 3.10 Reliability estimates: MSQ individual items

Item	Subscale	Corrected Item Total Correlation	Cronbach Alpha if Item Deleted
1	Planning	.47	.74
2	Planning	.41	.74
3	Planning	.20	.76
4	Planning	.20	.76
5	Planning	.15	.76
6	Considering the Audience	.23	.76
7	Considering the Audience	.39	.75
8	Considering the Audience	.35	.75
9	Monitoring	.36	.75
10	Monitoring	.33	.75
11	Monitoring	.27	.75
12	Monitoring	.10	.76
13	Monitoring	.57	.73
14	Monitoring	.56	.73
15	Monitoring	.41	.74

16	Monitoring	.11	.76
17	Evaluating	.19	.76
18	Evaluating	.28	.75
19	Evaluating	.44	.74
20	Evaluating	.49	.74
	Cronbach Alpha for the 20 items = .76		

The item by item analysis shows that the correlation coefficients of 11 items were above the .3 benchmark, and thus they contributed well to the overall reliability of the questionnaire. Items 13 and 14 grouped within the monitoring subscale were found to have the highest correlation coefficient (α = .57 and α = .56 respectively) with the overall set of items in the questionnaire. This was not surprising since these two items had also strong correlation coefficients (α = .54 and α = .61 respectively) with the monitoring subscale itself.

Accordingly, there are nine items (3, 4, 5, 6, 11, 12, 16, 17 and 18) with a correlation coefficient of below .3 when calculated with the overall questionnaire. These items had yielded positive correlations, albeit somewhat low for items 3, 6, 16, with the overall questionnaire in the pilot questionnaire. Since their reliability seemed to be adequate, the items were retained but revised for the questionnaire in Study 2. In particular, items 12 and 16 showed the lowest correlation coefficients with the total questionnaire (α = .10 and α = .11 respectively). Item 12 (*whether I was using the correct grammar*) was measuring the micro-level aspects of writing similar to items 10 and 11 and its score was expected to correlate with the rest of the questionnaire items but it did not. One possible reason for the weak correlation of item 16 (*how to connect different parts of my essay*) with the whole questionnaire may be the respondents' misinterpretation of the meaning of the word 'connect'. Interestingly, though, none of the items if deleted produced a Cronbach alpha that was significantly higher than .76.

In sum, after pilot testing the questionnaire in Study 1 and conducting item analysis, the clarity of the items and instructions, and the length of time necessary to complete the instrument was tried out prior to Study 2. Ideally, the revised questionnaire should have been once again pilot tested in a different context. However, given the limited number of participants and the insufficient time I was allowed to access them in the language school, there was no opportunity to do a final piloting.

3.6.3. Think-aloud protocols

Protocol data were analysed qualitatively according to the metacognitive strategies they reflected to show their use of strategies during the treatment. All audio-taped data, collected over the four sessions, were transcribed by the researcher. Each transcription was read several times in order to gain a thorough sense of the main ideas being expressed and to identify the themes and patterns during each reading. The defining criterion for determining what is metacognitive was those strategies that immediately preceded or followed a cognitive activity. I ensured that the statements demonstrating the exact generation of content in the actual essays were not included in the list of strategies, and only what the participants were trying to figure out prior to or following each chunk of composition was taken into consideration. This made the task rather difficult, because participants differed in their full or partial verbalisation of the content itself. Thus, identifying the strategy entailed checking each individual statement and matching the essays and transcriptions equally at the same time. The highlighted data were then coded and analysed manually.

The data analysis entailed some decisions regarding the steps to be taken to prepare the transcriptions. The participants' think-aloud verbalisations, having been transcribed using the ordinary writing system, were isolated depending on whether they corresponded to the actual texts written or referred to other writing operations and strategies (Manchón & Roca de Larios, 2007). The latter were the main focus of this study and were thus underlined as the relevant segments in the transcriptions. Again following Manchón and Roca de Larios' study, the instances of participants' repetitions or rereading were marked in italics and the revisions made to the texts while writing were signalled in the protocols.

A coding scheme adapted from Flavell, Miller and Miller's (2002) model of metacognition was devised. In their typology, metacognition is divided into metacognitive knowledge, and metacognitive monitoring and self-regulation. The former is further divided into knowledge about persons, tasks, and strategies, whereas the latter consists of application of strategies like planning, monitoring, and evaluating. A particular code was used for each strategy and then the absence/presence or the frequency of occurrence of each strategy was tallied in terms of descriptive counts and was eventually presented in a table.

A comparison was then made between the data from the think-aloud protocols and the list of metacognitive strategies taught during the four sessions.

In order to analyse the data, I first scored one protocol and assigned different codes for different categories. One issue which was initially a source of confusion was the distinction between cognitive and metacognitive strategies. As Schellings et al. (2012) have also observed, although the distinction between metacognition and cognition appears quite clear from a theoretical perspective, distinguishing different levels of cognition in the data at hand was not an easy job. One may argue that some of the strategies labelled as metacognitive are more of cognitive nature. However, as Meijer, Veenman, and van Hout-Wolters (2006, p. 218) have also claimed, some covert metacognitive execution activities are inferred from the more overt cognitive activities. For instance, re-reading a part of one's essay while writing is a cognitive activity in itself, but selecting only a particular part of the essay to re-read is undoubtedly metacognitive in nature, because that reflects a metacognitive decision for action.

3.7. Ethical Considerations

Many researchers emphasise the need to address ethical concerns in research in order to protect the rights of participants. In accordance with the ethical guidelines issued by the University of Auckland Human Participants Ethics Committee (UAHPEC), participants' privacy and data confidentiality were respected throughout the research process. I now outline the steps I took to enhance the ethical integrity of the study.

Firstly, prior to the beginning of the data collection process, ethics approval (Ref: 2010/318) was obtained (see Appendix G). An amendment to the ethics approval was also obtained on January 13, 2011, with regard to the change of location of the study from a university in Australia in Study 1 to a language institute in Iran in Study 2. Written consent forms were then obtained from participants, the head of the program of English instruction programs at the Australian university and the manager of the language institute in Iran prior to the commencement of the study. These consent forms served as a kind of contract for all parties, with the main aim of safeguarding participants' rights. They performed the task with only the researcher present throughout the treatment. Their teachers were not involved in the data collection process. They were reassured that the

results of the data collected were purely for research purposes and would not have anything to do with their performances in the courses they were undertaking (in Study 1) or the results of the placement test designed to assess the approximate level of students (in Study 2). No pressure was applied and it was made clear to participants their right to decline involvement without fear of any recriminations, and that they could withdraw at any time. Pseudonyms were used to protect the identity of participants and all data, including recordings and transcriptions which were labelled as per pseudonyms and kept securely. They were assured that all data would be anonymous in respect of any written report of the study and therefore not able to be traced back to any individual, and that data from each participant would not be shared with any other participant. In fact, the participants responded enthusiastically to the opportunity. Only two participants in Study 1 and three in Study 2 did not complete the intervention classes and many were interested in obtaining the results after the study was completed. Thus all parties involved were informed of the purpose of the study and what their participation would entail.

3.8. Evaluation of methodology

Based on the standards that applied linguistics researchers use in judging the soundness of their research, this section evaluates the methodology of the study, which is organised around the quantitative and qualitative methods utilised. I made a thorough attempt to combine both standards, by primarily stressing quantitative ones, but also by briefly addressing qualitative ones. These components include *reliability*, *replicability*, *validity*, and *generalisability* (Dörnyei, 2007).

3.8.1. Reliability

The standard of reliability is concerned with "the degree to which the results of a questionnaire, test, or other measuring instruments are consistent" (J. D. Brown, 2008, p. 492). In order to maximise reliability in this study, the research was carefully planned and designed from the beginning and the instruments and measures involved were precisely designed and piloted.

According to the IELTS organisation (2010), the reliability coefficient of writing module was calculated as .81–.89. In addition, the composite reliability estimate for the academic module of the test data from 2009 was .95 with a SEM of .24. One way of ensuring the

inter-rater reliability is the importance of specific marking criteria which should be as clear as possible and allow for appropriate gradations (Jonsson & Svingby, 2007). In IELTS rating criteria, there are four major categories, each with sub-definitions. Since July 2007, the scores for the writing module have been reported in whole or half-bands in the same way as Reading and Listening modules. This, however, does not mean the assessment criteria and the way examiners assess the performances have changed. Only the reporting of the scores on a more precise and informative scale is changed. This was done as a service to stakeholders, including recognising organisations, teachers, test takers, who require more information on a candidate's abilities. So, for instance, a 5.5 in Writing represents a very strong performance at the Band 5 level.

The written essays were administered repeatedly in the pre-test and post-test. Study 2 was a replica of Study 1 with the addition of a delayed post-test. In Study 2, a certified IELTS rater and I rated the written texts independently. Initially, there were instances that raters' understanding and using the criteria were not in the same way and did not allow common interpretation or mutual agreement. For example, there seemed to exist the possibility that the native-speaker co-rater tended to have got acclimatised to the linguistic errors of non-native users of language and thus was more tolerant of them. Thus, Bachman and Palmer's (1996, p. 222) six steps for rater training were followed in this study to increase the raters' mutual agreement on the criteria: (1) Read and discuss scales together. (2) Review language samples which have been previously rated by expert raters and discuss the ratings given. The samples to review were those of Study 1. (3) Practice rating a different set of language samples. Then compare the ratings with those of experienced raters. The ratings and how the criteria were applied in those samples were discussed by the two raters. (4) Rate additional language samples and discuss. (5) Each trainer rates the same set of samples. Check for the amount of time taken to rate and for consistency. (6) Select raters who are able to provide reliable and efficient ratings. The Spearman's correlation coefficient obtained was .77 which was statistically significant (r = .77, p (2-tailed) = .00).

In Study 2, a fellow doctoral student who was not involved in the research also rated 20% of the pre-post-test written texts I had rated, following the same detailed measures of accuracy and complexity as described aforementioned in Chapter 2. The co-rater was familiarised with the rating scheme from her own research. Discrepancies such as the total number of different grammatical verb forms as opposed to different grammatical verbs

were discussed and resolved. The Pearson product moment correlation coefficients for the scores of the two raters ranged from a high of .93 for error-free verb forms to a low of .86 for syntactic complexity.

In order to enhance the consistency of protocol findings, a code-recode procedure was conducted on the data during the analysis phase of the study. After coding a set of think-aloud data for instance, I waited for a month and then returned and recoded a subset of the data to compare the results and check for intra-coder reliability. The next step taken was that a coded protocol was taken as a sample and was discussed with my supervisor. She then independently scored a second protocol and the discrepancies in assigning the codes were discussed afterwards. Finally, once the inter-rater reliability was established (r =.71), I myself continued coding the remainder of the protocols.

3.8.2. Replicability

Replicability is a standard in research which concerns the degree to which researchers can replicate or repeat the study in a new context according to the provided information by previous studies. The objective is to carry out the research as it was originally conducted (J. D. Brown, 2008, p. 492). In the methodology chapter, the characteristics and selection of participants in the study were completely described. Similarly, the instruments as well as the procedures followed in collecting the data, scoring or coding the instruments, and analysing the results were presented in detail. During the span of my doctoral candidature, the methodology and results of the two studies were presented in departmental seminars and professional conferences to fully reveal the design and analysis of the study on which interpretations were made. Also, in order to examine, confirm, reject, or modify the original coding scheme of think-aloud protocols, three samples were double-coded with my supervisor and other fellow doctoral students in a departmental seminar.

3.8.3. Validity

The standard of validity in research refers to “the correctness and appropriateness of the interpretations that a researcher makes of his/her study” (Gass, 2010, p. 12). When planning and conducting this project, the internal validity of this study was enhanced by triangulation which, as mentioned earlier, was employed in various ways. In terms of using multiple and different methods for data collection, the study integrated individual

participants' audio-recorded think-alouds and questionnaires along with their written essays. Data were collected in an extensive manner from six weeks in Study 1 to over nine weeks in Study 2. Triangulation also took place through data analysis and interpretation of findings to maximise the accuracy and completeness of the research results. In order to measure participants' written texts with confidence and certainty, multiple measures were used to reflect their written ability (construct validity). More importantly, the overall measure of writing proficiency was used in much the same way as the well-established writing module of IELTS (criterion-related validity).

3.8.4. Generalisability

The standard of generalisability signifies the quality in research which makes it capable of generalisation to other similar groups. According to this standard, the results of a study can be justifiably applied to a larger population (J. D. Brown, 2008, p. 493). In order to improve the generalisability of the study, careful attention was paid to use random sampling of participants to be representative of the population to which the study wished to generalise. In this randomisation procedure, each participant had an equal and independent chance of being assigned to any of the groups and then random assignment of groups to treatment conditions of the study (experimental or control) occurred. However, no assumptions were made at the beginning of the project to apply interpretations of the qualitative data analysis to other contexts. The findings were descriptive in nature and were merely attempted to represent individual participants' perspectives of strategies in essay writing. However, detailed descriptions of theoretical and methodological orientations were provided for other researchers to determine if contextual similarities were sufficient to be applicable in their own research.

Summary

In this section, it was ensured that collecting a significant number of quantitative and qualitative data using complementary instruments, and thorough analyses of the data collected were offered in support of the validity and reliability of the study.

Conclusion

The research questions which framed this study called for a methodology which would have both a holistic orientation toward L2 learners' written products and a detailed

description of what actually occurs in the process of writing. This chapter has presented information about the research design, context, participants, research instruments, data collection methods and analysis procedures for this project. Further, the modifications that were made to the design of the study after Study 1 have been outlined. The overall design of the study aimed at answering the four research questions. In the subsequent chapters, the results of the data gathered from writing essays, questionnaires, and think-aloud protocols for each of the two studies will be presented and evaluated.

CHAPTER 4. RESULTS AND DISCUSSION OF STUDY 1: RESEARCH QUESTIONS 1-3

In order to gain a more complete picture of the complex nature of writing, this study utilised a mixed-method research in respect of the data collection instruments and the analysis of the data collected. The purpose of this chapter is to report the results of the quantitative data analysis in Study 1 in order to provide some answers to the following three research questions:

1. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *general writing proficiency* of L2 learners' written products?
2. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *accuracy* of L2 learners' written products?
3. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *complexity* of L2 learners' written products?

4.1. Results

In order to compare participant groups and measure the degree of within-group changes as a result of different treatments, a pre-test and a post-test were conducted. The normal distribution of the four groups' scores on all variables, which were analysed based on a holistic rating (i.e. IELTS rating scale) and a detailed rating (i.e. five variables of accuracy and complexity), was tested in terms of skewness and kurtosis. For those normally distributed, the results of the pre-test and post-test written productions of each group were then compared using a series of ANOVA to find probable significant changes, followed by post-hoc Tukey tests to explore possible differences in pairs of means of the groups on their written tasks. For significantly non-normal data, the Kruskal-Wallis test was used, followed by a Mann-Whitney U test. The alpha value for statistical significance was set at .05. The results which did not show statistical significance are provided in Appendix F.

As mentioned earlier in chapter 3, a sample of the IELTS second writing task was administered in order to ascertain the homogeneity of participants in the study. Overall, the mean scores of the participants' written texts equalled 5.21 (*SD* = .40). The results of the Kruskal-Wallis test ensured that there were no statistically significant differences between the four groups before embarking upon the whole treatment ($H(3) = .45$, $p = .92$). Thirty five participants were then randomly assigned into four groups. The pre-test and post-test writing tasks, once collected, were analysed based on IELTS rating scale and two measures used extensively in task-based research: accuracy (error-free clauses, and error-free verb forms) and complexity (syntactic complexity, syntactic variety, and lexical complexity). This research approach is text-oriented, and views writing as "a product, an artefact of activity which can be studied independently of users by counting features and inferring rules" (Hyland, 2010, p. 191). In the following section, the results obtained from the analyses of general writing proficiency, accuracy and complexity in the pre-test and post-test writing are discussed separately for each measure.

4.1.1. Research Question 1

Effects of metacognitive strategy instruction and pre-task planning on the general writing proficiency of L2 learners' written products

The writing tasks were analysed holistically to determine whether pre-task planning and metacognitive instruction influenced participants' overall quality of L2 writing production. This holistic rating was based on IELTS detailed descriptive criteria and rating scales. The descriptive results of pre-test and post-test writing tasks were first individually measured (see Table 4.1). In both tasks, the scores were almost equally clustered. In the pre-test, all groups appeared to have performed homogenously. The range of lowest and highest scores were in the non-metacognitive groups; from the lowest of *M* = 4.92, *SD* = .78 for Group -M-P to the highest of *M* = 5.25, *SD* = 1.00 for Group -M+P. In the post-test, the mean score ranges were from *M* = 5.28, *SD* = .80 for Group -M-P to *M* = 6.00, *SD* = .62 for Group +M+P. Further analysis was required to show whether the mean scores of groups significantly increased from the pre-test to the post-test.

Table 4.1 Descriptive statistics: General writing proficiency

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	10	5.20	.78	.24
	+M-P	10	5.00	.91	.28
	-M+P	8	5.25	1.00	.35
	-M-P	7	4.92	.78	.29
	Total	35	5.10	.84	.14
Post-test	+M+P	10	6.00	.62	.19
	+M-P	10	5.55	.59	.18
	-M+P	8	5.62	1.32	.47
	-M-P	7	5.28	.80	.30
	Total	35	5.64	.86	.14

Note. M: Metacognitive instruction, P: Planning

In order to specify whether any of the groups outperformed the other, either in the pre-test or post-test, the Kruskal-Wallis test was used, in view of the fact that the distribution of scores was non-normal. The results showed that none of the groups differed with regard to their performance in either pre-test (H (3) = .90, p = .82) or post-test (H (3) = 4.45, p = .21).

In order to ascertain whether the changes exhibited by groups in response to the treatment over time were statistically significant, the Wilcoxon signed-rank test, the non-parametric counterpart of dependent t-test, was used (see Table 4.2).

Table 4.2 Wilcoxon signed-rank test: Pre-test to post-test progress in general writing proficiency

	+M+P	+M-P	-M+P	-M-P
Z*	-2.40	-1.26	-1.38	-1.89
Asymp. Sig. (2-tailed)	.01*	.20	.16	.06

Note. * Based on negative ranks

*. The mean difference is significant at the .05 level.

According to the results, only Group +M+P made significant progress from the pre-test to the post-test (p = .01, d = .53). The effect size represented a large change in this group's level of writing proficiency (i.e. it is above .5, the threshold of Cohen's benchmark of large effect size). Differences in other groups' performances from the pre-test to the post-test did not uncover significance.

Summary

No significant between-group differences were found in terms of general writing proficiency. The analysis for progress over time indicated that Group +M+P created significantly higher-quality written content in the post-test, compared to the pre-test.

4.1.2. Research Question 2

Effects of metacognitive strategy instruction and pre-task planning on the accuracy of L2 learners' written products

In this section, the results are discussed separately for each two measures of accuracy, i.e. error-free clauses and error-free verb forms.

4.1.2.1. Accuracy: Error-free clauses

Descriptive statistics for error-free clauses in pre-test and post-test are displayed in Table 4.3. All in all, the mean scores of the four groups tended to improve over time for this measure of accuracy. However, the groups' mean scores were closely clustered in both task performances. In the pre-test, the four groups ranged from 38% error-free clauses (Group -M+P) to 44% error-free clauses (Group +M+P). In the post-test production, the range was from Group +M-P producing 51% error-free clauses to Group -M+P and Group -M-P producing 55% error-free clauses.

Table 4.3 Descriptive statistics: Error-free clauses

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	10	.44	.09	.03
	+M-P	10	.42	.14	.04
	-M+P	8	.38	.11	.03
	-M-P	7	.43	.13	.05
	Total	35	.42	.12	.02
Post-test	+M+P	10	.52	.08	.02
	+M-P	10	.51	.13	.04
	-M+P	8	.55	.06	.02
	-M-P	7	.55	.11	.04
	Total	35	.53	.10	.01

Note. M: Metacognitive instruction, P: Planning

Data were then submitted to one-way ANOVA analysis separately for the pre-test and post-test to determine whether there were any significant differences between groups. Differences in planning time conditions and metacognitive instructions among four groups did not lead to significant differences in the global accuracy rating of error-free clauses throughout the treatment.

A series of paired-samples t-test was then calculated to show whether the differences between the means of the two conditions for each group was sufficiently large. The mean difference between scores is the difference between the mean scores of the pre-test and post-test conditions. As shown in Table 4.4, the fact that the t-value was a negative number revealed that the pre-test had a smaller mean than the post-test. In fact, all treatment groups, except Group +M-P, produced significantly more error-free clauses over time, with $t(9) = -3.11$, $p = .01$, $d = .72$ for Group +M+P; $t(7) = -4.56$, $p = .00$, $d = .86$ for Group -M+P; and $t(6) = -3.74$, $p = .01$, $d = .83$ for Group -M-P. As reported, these improvements represented very large effect sizes. Therefore, as well as being statistically significant, these large effect sizes represent a substantive finding.

Table 4.4 Paired-sample t-test: Error-free clauses

		Paired Differences					
		Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
+M+P	Pre-test - Post-test	-.08*	.08	.025	-3.11	9	.01
+M-P	Pre-test - Post-test	-.08	.17	.055	-1.51	9	.16
-M+P	Pre-test - Post-test	-.16*	.10	.037	-4.56	7	.00
-M-P	Pre-test - Post-test	-.11*	.08	.031	-3.74	6	.01

Note. M: Metacognitive instruction, P: Planning

*. The mean difference is significant at the .05 level.

To compare the changes exhibited by the four groups in response to the treatment, the gain scores, or the post-test minus pre-test score, were computed. A one-way ANOVA was then used to compare the mean difference scores of groups. The results revealed that the pre-to-post-test variance and the measurement of change observed from paired sample t-test for Groups +M+P, -M+P and -M-P did not differ significantly among groups ($p = .60$), leading us to the conclusion that all three groups significantly progressed, but at almost the same rate.

4.1.2.2. Accuracy: Error-free verb forms

As displayed in Table 4.5, there were clear differences between groups in terms of error-free verb forms in the pre-test. Groups +M+P and -M-P produced 71% error-free verbs, and Group +M-P produced 63%, while Group -M+P only produced 57% error-free verbs. As for the post-test, the range of mean scores was relatively narrow, with Group +M+P still producing the highest of 72% error-free verbs, while Group -M-P decreased to the lowest of 58%.

Table 4.5 Descriptive statistics: Error-free verb forms

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	10	.71	.11	.03
	+M-P	10	.63	.09	.03
	-M+P	8	.57	.08	.03
	-M-P	7	.71	.08	.03
	Total	35	.65	.11	.01
Post-test	+M+P	10	.72	.07	.02
	+M-P	10	.65	.07	.02
	-M+P	8	.60	.06	.02
	-M-P	7	.58	.19	.07
	Total	35	.64	.11	.01

Note. M: Metacognitive instruction, P: Planning

According to ANOVA results of pre-test production, there was a significant difference between groups in the provision of error-free verbs with a large effect size, even before starting the unit of instruction (p = .02, d = .51). Table 4.6 displays the results of this analysis in the pre-test.

Table 4.6 ANOVA: Error-free verb forms (Pre-test)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.11*	3	.03	3.77	.02
Within Groups	.31	31	.01		
Total	.42	34			

*. The mean difference is significant at the .05 level.

In order to specify where the differences in the pre-test lay, post-hoc Tukey test analysis (Table 4.7) indicated that Group +M+P and Group -M-P produced writing with a significantly higher proportion of target-like verbs than Group -M+P (p = .03, d = .58; p = .04, d = .65 respectively).

Table 4.7 Tukey: Error-free verb forms (Pre-test)

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
+M+P	+M-P	.07	.04	.36
	-M+P	.13*	.04	.03
	-M-P	-.00	.04	1.00
+M-P	+M+P	-.07	.04	.36
	-M+P	.06	.04	.53
	-M-P	-.07	.04	.40
-M+P	+M+P	-.13*	.04	.03
	+M-P	-.06	.04	.53
	-M-P	-.14*	.05	.04
-M-P	+M+P	.00	.04	1.00
	+M-P	.07	.04	.40
	-M+P	.14*	.05	.04

Note. M: Metacognitive instruction, P: Planning
*. The mean difference is significant at the .05 level.

As for the post-test production, the differences between groups were minimal. As shown in Table 4.8, although the gain was not big enough to be justified statistically, it approached significance and represented a medium-sized effect (p = .051, d = .46). Post-hoc Tukey test analysis only showed a non-significant trend for Group +M+P to produce more accurate verbs than Group -M-P (p = .06).

Table 4.8 ANOVA: Error-free verb forms (Post-test)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.09	3	.03	2.89	.051
Within Groups	.35	31	.01		
Total	.44	34			

*. The mean difference is significant at the .05 level.

In order to ascertain whether or not the two sets of scores obtained in the pre-test and post-test for the same set of subjects within each group differed from each other, a series of paired sample t-tests were conducted. Unlike other groups, the *t*-value for Group -M-P was a positive number ($t(6) = 1.84$), which indicates that the pre-test condition had even a bigger mean than the post-test condition. However, none of the differences in any of the groups was significant ($p > .05$).

Summary

With respect to accuracy, no paired comparisons reached statistical significance where the percentage of clauses without errors was concerned. Groups +M+P, -M+P and -M-P all wrote significantly more accurately in respect of their use of clauses, when progress over time was measured. A different picture emerged, however, with regard to the percentage of accurately used verbs in the pre-test, with Groups +M+P and -M-P displaying higher rates than Group -M+P. These differences disappeared in the post-test. However, none of the four groups manifested any progress over time.

4.1.3. Research Question 3

Effects of metacognitive strategy instruction and pre-task planning on the complexity of L2 learners' written products

Three variables were assessed to measure the complexity of language used in the participants' written essays: syntactic complexity, syntactic variety, and lexical variety. The third research question pertains to the analysis of these three measures.

4.1.3.1. Complexity: Syntactic complexity

As displayed in Table 4.9, in terms of syntactic complexity, the mean scores of the four groups were closely clustered in both task performances. In the pre-test, they ranged from Group +M-P producing the highest of $M = 2.06$, $SD = .42$, with Groups -M+P and -M-P following it almost equally, and Group +M+P producing the lowest of $M = 1.74$, $SD = .35$. As for the post-test, a slight downward trend was recorded for Group +M-P and a small upward trend was observed for the other three groups. The results of the one-way ANOVA confirmed that these slight differences among groups in the pre-test and post-test were non-significant in the syntactic complexity of participants' production, as measured by clauses per T-unit ($p > .05$).

Table 4.9 Descriptive statistics: Syntactic complexity

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	10	1.74	.35	.11
	+M-P	10	2.06	.42	.13
	-M+P	8	1.92	.28	.10
	-M-P	7	1.98	.54	.20
	Total	35	1.92	.40	.06
Post-test	+M+P	10	1.87	.34	.10
	+M-P	10	2.00	.53	.16
	-M+P	8	2.08	.40	.14
	-M-P	7	1.99	.41	.15
	Total	35	1.98	.41	.07

Note. M: Metacognitive instruction, P: Planning

The results of the paired sample t-test, as illustrated in Table 4.10, showed that although the mean score of Group +M-P's performance on the pre-test seemed to have been even more syntactically complex than the post-test (i.e. this is shown by the positive t-score gained in the paired sample t-test ($t(9) = 1.17$), neither this group, nor other groups achieved any significant progress from the pre-test to the post-test. This finding leads us to the conclusion that neither of the two treatments had any considerable impact on the improvement of syntactic complexity in the texts produced by these participants.

Table 4.10 Paired-sample t-test: Syntactic complexity

		Paired Differences					
		Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
+M+P	Pre-test - Post-test	-.12	.48	.15	-.84	9	.42
+M-P	Pre-test - Post-test	.05	.15	.04	1.17	9	.27
-M+P	Pre-test - Post-test	-.15	.29	.10	-1.51	7	.17
-M-P	Pre-test - Post-test	-.00	.28	.10	-.07	6	.94

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

4.1.3.2. Complexity: Syntactic variety

According to the descriptive statistics displayed in Table 4.11, the mean scores of groups were closely clustered in the pre-test, showing that they produced writing at a similar level

of syntactic variety, as measured by variety in verb constructions. With regard to the post-test production, Group +M+P produced the most syntactically varied writing (M = 9.10, SD = 1.37). Groups +M-P and -M+P ranked the second and third, and Group -M-P the last (M = 7.43, SD = 1.27).

Table 4.11 Descriptive statistics: Syntactic variety

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	10	6.90	.99	.31
	+M-P	10	8.60	1.35	.42
	-M+P	8	7.88	1.80	.63
	-M-P	7	7.71	1.79	.68
	Total	35	7.77	1.55	.26
Post-test	+M+P	10	9.10	1.37	.43
	+M-P	10	8.80	1.03	.32
	-M+P	8	8.13	1.12	.39
	-M-P	7	7.43	1.27	.48
	Total	35	8.46	1.31	.22

Note. M: Metacognitive instruction, P: Planning

Significant differences were found consistently among groups in terms of syntactic variety of production in their post-test writing (p = .04, d = .48), but not in their pre-test (Table 4.12).

Table 4.12 ANOVA: Syntactic variety

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	14.56	3	4.85	2.22	.10
	Within Groups	67.60	31	2.18		
	Total	82.17	34			
Post-test	Between Groups	13.59*	3	4.53	3.11	.04
	Within Groups	45.08	31	1.45		
	Total	58.68	34			

Note. M: Metacognitive instruction, P: Planning

*. The mean difference is significant at the .05 level.

Results of post-hoc Tukey test analysis are shown in Table 4.13. They indicate that only Group +M+P produced significantly more syntactically varied writing than Group -M-P (p

= .04, *d* = .53). None of the other between-group comparisons showed any significance in the post-test.

Table 4.13 Tukey: Syntactic variety (Post-test)

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
+M+P	+M-P	.30	.53	.94
	-M+P	.97	.57	.33
	-M-P	1.67*	.59	.04
+M-P	+M+P	-.30	.53	.94
	-M+P	.67	.57	.64
	-M-P	1.37	.59	.11
-M+P	+M+P	-.97	.57	.33
	+M-P	-.67	.57	.64
	-M-P	.69	.62	.68
-M-P	+M+P	-1.67*	.59	.04
	+M-P	-1.37	.59	.11
	-M+P	-.69	.62	.68

Note. M: Metacognitive instruction, P: Planning
*. The mean difference is significant at the .05 level.

As shown in Table 4.14, the four series of paired sample t-test show that only Group +M+P showed a significant improvement from the pre-test to the post-test in this measure of complexity; $t(9) = -3.71$, $p = .00$, $d = .77$. Although the means scores of Groups +M-P and -M+P also showed an increase, they did not return statistically significant results. The positive t-score for Group -M-P ($t(6) = .79$) shows a decline in performance from the pre-test to the post-test, however again the decrease was not significant ($p = .45$).

Table 4.14 Paired-sample t-test: Syntactic variety

		Paired Differences					
		Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
+M+P	Pre-test - Post-test	-2.20*	1.87	.59	-3.71	9	.00
+M-P	Pre-test - Post-test	-.20	1.03	.32	-.61	9	.55
-M+P	Pre-test - Post-test	-.25	1.16	.41	-.60	7	.56
-M-P	Pre-test - Post-test	.28	.95	.36	.79	6	.45

Note. M: Metacognitive instruction, P: Planning
*. The mean difference is significant at the .05 level.

4.1.3.3. Complexity: Lexical variety

The descriptive results presented in Table 4.15 demonstrate that in both pre-test and post-test, the four groups were closely clustered in terms of their production of lexically varied writing, as measured by Giraud index of lexical richness.

Table 4.15 Descriptive statistics: Lexical variety

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	10	7.00	.62	.19
	+M-P	10	7.78	.58	.18
	-M+P	8	7.11	.96	.34
	-M-P	7	7.45	.65	.24
	Total	35	7.34	.75	.12
Post-test	+M+P	10	7.69	.71	.22
	+M-P	10	7.67	.84	.26
	-M+P	8	7.25	.83	.29
	-M-P	7	7.44	.54	.20
	Total	35	7.53	.74	.12

Note. M: Metacognitive instruction, P: Planning

The results of the ANOVA revealed that there were no statistically significant differences among the performances of groups in the pre-test or post-test ($p = .09 > .05$, $p = .57 > .05$ respectively). Although the statistical analysis did not uncover a significant result for the pre-test ($p = .09$), a marginal trend was found for lexical variety with the partial d of .46 indicating a medium effect size.

According to the paired sample t-tests, as presented in Table 4.16 below, although the mean scores of Group -M+P showed a modest increase from the pre-test to the post-test and Groups +M-P and -M-P showed a small decrease, none of these over-time comparisons uncovered significant results. Only Group +M+P obtained a significant improvement in lexical complexity from the pre-test to the post-test; $t(9) = -2.60$, $p = .03$, $d = .65$.

Table 4.16 Paired-sample t-test: Lexical variety

		Paired Differences					
		Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
+M+P	Pre-test - Post-test	-.68*	.83	.26	-2.60	9	.03
+M-P	Pre-test - Post-test	.10	.86	.27	.37	9	.72
-M+P	Pre-test - Post-test	-.13	.72	.25	-.54	7	.60
-M-P	Pre-test - Post-test	.00	.69	.26	.02	6	.97

Note. M: Metacognitive instruction, P: Planning

*. The mean difference is significant at the .05 level.

Summary

Where complexity was concerned, all groups wrote equally complex argumentative essays in the pre-test and post-test. Only with regard to the total number of different grammatical verb forms used in the task, did Group +M+P outscore Group -M-P in the post-test. Although none of the groups manifested any progress rate in terms of the syntactic complexity of language use, Group +M+P progressed with respect to both syntactic and lexical variety.

4.1.4. Summary of results

A summary of results is presented in Tables 4.17 and 4.18. All groups were equivalent in all measures at the outset, with the exception of the production of error-free verbs, where Groups +M+P and -M-P produced significantly higher percentages of error-free verbs compared to Group -M+P. However, these differences disappeared in the post-test. In the post-test production, all groups performed almost identically in all measures, with one exception, namely the fact that Group +M+P produced greater syntactic variety than Group -M-P.

Table 4.17 Summary of between-group results in Study 1

Measures		Pre-test	Post-test
General writing proficiency		n.s.	n.s.
Accuracy	Error-free clauses	n.s.	n.s.
	Error-free verbs	+M+P > -M+P -M-P > -M+P	n.s.
Complexity	Syntactic complexity	n.s.	n.s.
	Syntactic variety	n.s.	+M+P > -M-P
	Lexical variety	n.s.	n.s.

In terms of gain scores, Groups +M+P, -M+P and -M-P were all significantly more accurate in the use of clauses in their writing. Group +M+P also made noticeable gains in their general writing proficiency, syntactic variety and lexical variety. Yet no groups produced syntactically more complex texts, containing error-free verb forms in the post-test compared to the pre-test.

Table 4.18 Summary of within-group results in Study 1

Measures		Gain scores (Pre-Post)
General Writing Proficiency		+M+P ↑
Accuracy	Error-free clauses	+M+P ↑, -M+P ↑, -M-P ↑
	Error-free verbs	n.s.
Complexity	Syntactic complexity	n.s.
	Syntactic variety	+M+P ↑
	Lexical variety	+M+P ↑

Note. ↑ shows significant increase.

4.2. Discussion of Results

Study 1 sought to examine the different effects of pre-task planning and metacognitive instruction on the overall writing proficiency, accuracy and complexity of ESL learners' written text products. Participants were randomly assigned into four different conditions and produced argumentative writing in pre- and post-tests. Their written texts were then examined to determine whether different treatments could possibly influence the quality of participants' writing products differently. This section provides a detailed discussion concerning different measures of writing in response to the following research questions:

1. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *general writing proficiency* of L2 learners' written products?
2. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *accuracy* of L2 learners' written products?
3. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *complexity* of L2 learners' written products?

These research questions reflect two key means to measure learner production, i.e. either by external ratings, which are mainly preferred by language testers, or discourse analytic measures, which are preferred by SLA researchers (Ellis, 2005, p. 30). In general, the results reported earlier in this chapter indicated that these two types of holistic and analytic measures would reveal different effects on written performance. What follows is the discussion of the main findings concerning the three aspects of performance addressed in the research questions: general writing proficiency, accuracy, and complexity.

4.2.1. Research Question 1

Effects of metacognitive strategy instruction and pre-task planning on the general writing proficiency of L2 learners' written products

The first research question concerns the effects of metacognitive instruction and pre-task planning on the general writing proficiency of L2 learners' written products. Participants' writing proficiency was measured by means of external rating of IELTS which specifies levels of performance for target competency in linguistic terms described holistically across nine bands (see Appendices D & E). Although the means of the IELTS scores in the planning groups (+M+P and -M+P) were comparatively higher than those of the non-planning groups (+M-P and -M-P) both in the pre-test and post-test, the comparison for this measure among four groups did not achieve statistical significance. Based on the results gained from this proficiency test given before and after the treatment, only Group +M+P showed significant progress from the pre-test to the post-test (see Figure 4.1). Therefore, the answer to this research question is that metacognitive instruction when combined with pre-task planning produced significant progress over time in the general

writing proficiency of learners although between-group comparisons did not uncover any significant results.

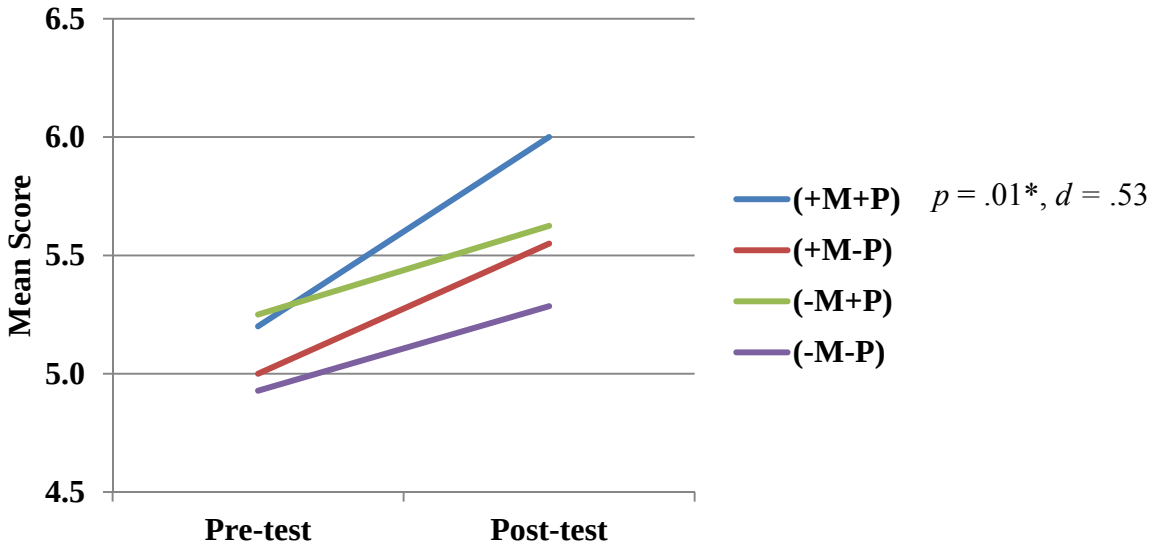


Figure 4.1 General writing proficiency

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

Although the study participants were not fully aware of what was expected of them during the first experimental session, they came to the subsequent sessions with prior experience and familiarity with the type of task they were required to perform. Thus one would expect a source of variation, at least due to the practice effect. To explain this in terms of Anderson's (J. R. Anderson, 1983) procedural and declarative knowledge representations, participants who were confronted with the task for the first time were expected to have had to access a control mechanism that could help them draw on explicit stored declarative knowledge. Accordingly, participants, now possessing some procedural linguistic and rhetorical skills, would perform differently in the subsequent task because of the familiarity with the experimental situation or the measures being manipulated. Only Group +M+P, however, showed progress from the pre-test to the post-test. Kellogg's (1996) model of writing characterises three basic systems involved in written text production: formulation, execution and monitoring. It is possible to say that when time allowed for planning, Group +M+P, who were not only advantaged by the provision of pre-task planning, but were also instructed how to perform using metacognitive strategies, acquired

ample knowledge during the course of the treatment and then applied it in the post-test. This knowledge helped them to formulate written plans, to execute the actual production of sentences, and to monitor their execution of texts.

Other between-group and within-group comparisons for this dependent variable of general writing proficiency did not produce significant results. The results corroborate those of Dellerman, Coirier, and Marchand's (1996), Dujsik's (2008), Ojima's (2006), and Shin's (2008) studies which found no significant differences between the planning and no planning conditions regarding the overall quality of writing tasks. It is possible that the planning and/or metacognitive conditions enhanced the quality of writing in ways not measured by holistic scores of writing. For this reason, the participants' written text products were also evaluated by means of discourse analytic measures where different dimensions of their performance – accuracy and complexity – were rated. The next two sections deal with discussion of the second and third research questions regarding the two dimensions of accuracy and complexity to explore whether or not they uncovered advantages for pre-task planning and/or metacognitive instruction.

4.2.2. Research Question 2

Effects of metacognitive strategy instruction and pre-task planning on the accuracy of L2 learners' written products

It should be noted that, as illustrated in Figure 4.2, no significant differences were detected among groups when the error-free clauses of their writing were measured in either the pre-test or post-test. Only when the progress from the pre-test to the post-test was taken into account, Groups +M+P, -M+P and -M-P showed significant improvements in their production of written texts containing error-free clauses, a result which was not statistically significant between groups when examined separately for pre-test and post-test. Therefore, the answer to this research question with regard to the first measure of accuracy is negative. While both Group +M+P and Group -M+P had the opportunity to specifically concentrate on pre-task planning, this did not exclusively contribute to their more skilled performance than other non-planning groups in their production of accurate clauses in any significant way. No superior performance was observed between the metacognitively instructed and non-metacognitively instructed groups.

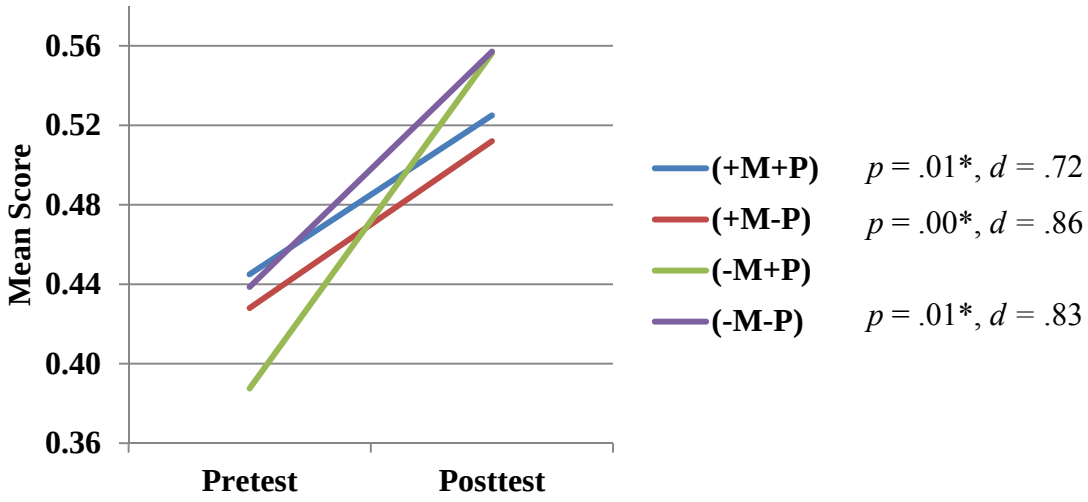
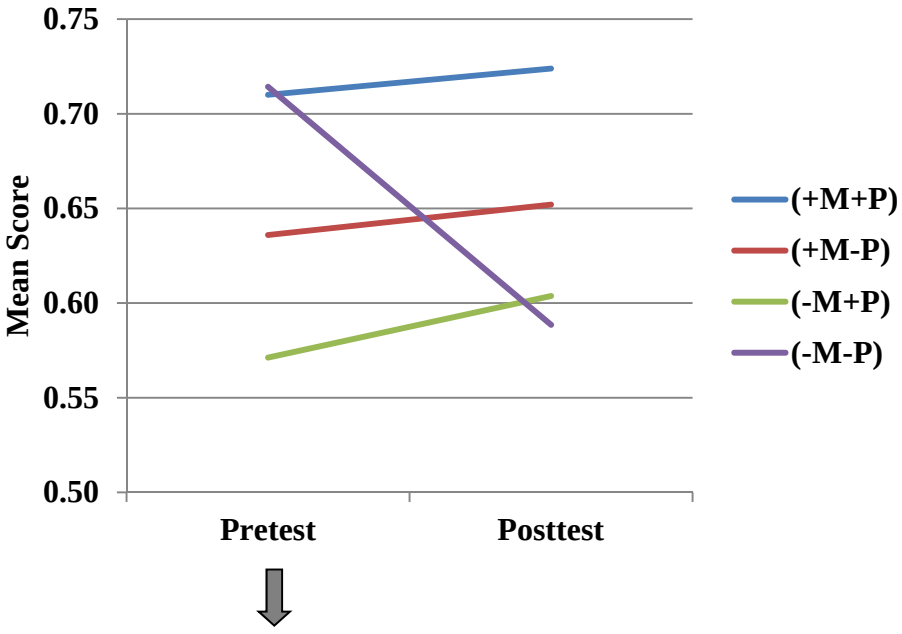


Figure 4.2 Accuracy (Error-free clauses)

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

In case of error-free verbs, the overall picture was more complicated. As demonstrated in Figure 4.3, there already existed differences in the pre-test, with Groups +M+P and -M-P both outperforming Group -M+P. The superior performance of Group +M+P than Group -M-P approached statistical significance in the post-test (p = .06), however, gain scores, i.e. comparing any one group with their own pre- and post-tests, showed no significant differences between groups on their performance from the pre- to post-test, suggesting no advantage for pre-task planning. Thus, this tendency toward greater accuracy of Group +M+P than Group -M-P could not be explained as a result of the different treatments they received. This was because the examination of gain scores did not uncover significant results. The treatment could therefore not help the metacognitive (with or without planning) groups to compensate for the probable processing limitations in attending to linguistic content during the task performance, and thus maximising accuracy. Thus, the answer to the second research question, with regard to both measures of accuracy, is negative.



+M+P > -M+P (p = .03*), d = .58
-M-P > -M+P (p = .04*), d = .65

Figure 4.3 Accuracy (Error-free verbs)

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

All in all, the results of the present study showed that neither pre-task planning, nor metacognitive instruction had any large effect on accuracy. Previous studies of pre-task planning have produced mixed results where accuracy is concerned, and no documented research has ever been reported on the impact of metacognitive instruction on accuracy. The findings in the current study accord with the results reported in a number of previous task-based research studies, which also failed to demonstrate any effects of pre-task planning on accuracy (e.g. Ojima, 2006; Yuan & Ellis, 2003). One possible explanation for the lack of progress in the measures of accuracy can be the nature of writing tasks. Due to their cognitive overload, argumentative tasks are demanding in nature, with their emphasis on the grammatical, lexical, structural, and rhetorical features of the argumentative genre. They impose a complex set of processes to be carried out in a limited capacity working memory. This likely encouraged greater attention to the propositional and organisational content of the argumentation, rather than to the accuracy of the clauses or verb forms. Skehan's (1998) dual-processing cognitive model suggests that the learner's rule-based system requires more time and attentional capacity to access than the exemplar-based system which contains easily-accessed lexically stored knowledge. In line

with this model, the failure of the results to show significant differences for accuracy measures may be also due to the fixed limited time the participants had to fully access their explicit rule-based grammatical knowledge during their task performance. After all, as claimed by Wendel (1997), accuracy occurs as a consequence of the moment-by-moment decisions that learners make during task performance. Within the pressure of real-time tasks, the adjustments and improvements required for the effective control of accurate language use may not have been truly feasible, even for the groups who were provided with the planning opportunities.

Another possible explanation is that, given the general nature of metacognitive strategies, one would not expect a dramatic change at the syntactic level of performance. Furthermore, the strategy or approach taken by the teacher or programme has important consequences for the learning outcomes of students. In accordance with the instructions given to the four groups, the study did not aim to instruct and reinforce the error-free usage of language with no attempt to draw learners' explicit attention to linguistic form. In other words, teaching techniques and procedures did not take into account the structural and developmental errors that reflect the learners' competence at a particular stage of interlanguage development and illustrate some of the general characteristics of language acquisition. What could be optimally expected and observed, particularly in case of error-free clauses, was the role metacognitive strategies might play in helping learners to automatise the grammatical knowledge they had acquired prior to or during the treatment and naturally present the target features when the tasks were performed. This most probably explains why Groups -M+P, -M-P and -M-P progressed in their production of texts containing comparatively more accurate clauses.

4.2.3. Research Question 3

Effects of metacognitive strategy instruction and pre-task planning on the complexity of L2 learners' written products

To answer the third research question, three measures of complexity were investigated: syntactic complexity, syntactic variety, and lexical variety. Comparing the differences among groups with regard to complexity, the present study found no statistically significant differences among groups on syntactic complexity at the pre-test and post-test, or even when the progress of pre-to-post-test was measured (see Figure 4.4). In other

words, all groups were equivalent in pre- and post-test performances on syntactic complexity. So the answer to this measure of complexity in the research question is negative.

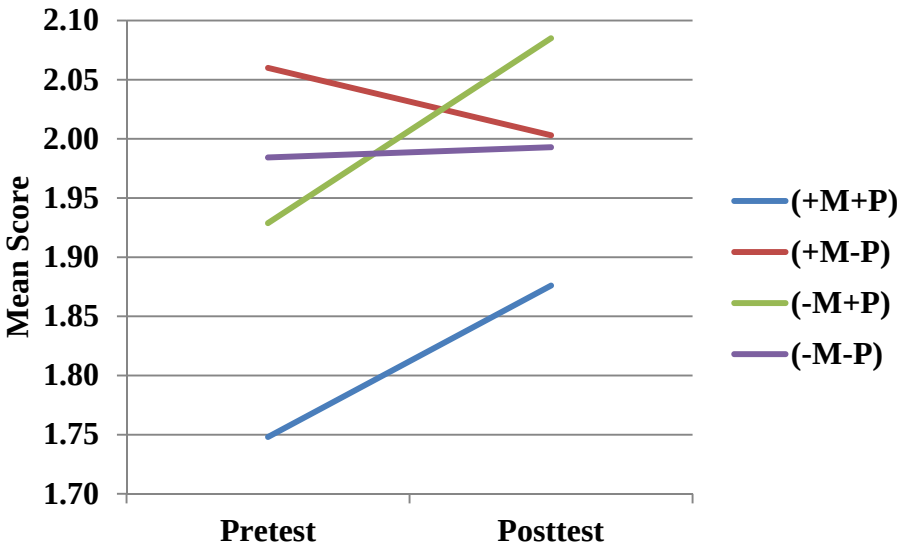


Figure 4.4 Complexity (Syntactic complexity)

Note. M: Metacognitive instruction, P: Planning

The results showed a statistically significant difference between groups for syntactic variety. More specifically, Group +M+P, given planning and metacognitive instruction, produced significantly more syntactically varied writing in the post-test than Group -M-P who received neither planning nor instruction (p = .04) (see Figure 4.5). When the over-time progress was measured, Group +M+P showed an increase from the pre-test to the post-test. This suggests that they used the pre-task planning to focus their attention on the propositional content and organisation of the main arguments of their writing, but at the same time to search for specific linguistic means, for instance various verb forms, to encode their meaning while writing. Groups +M-P and -M+P, however, showed lack of significant results for syntactic variety over time. This might suggest that planning alone (Group -M+P) or instruction alone (Group +M-P) had no effect on syntactic variety in written production.

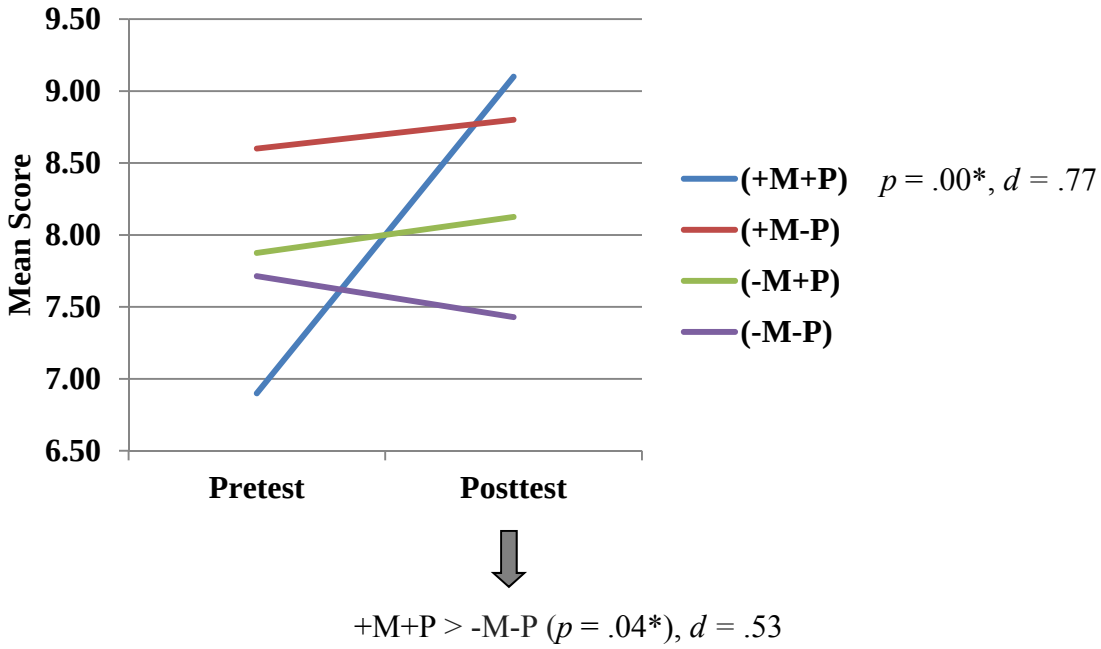


Figure 4.5 Complexity (Syntactic variety)

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

The results of the study indicated that no between-group differences were evident for the lexical variety in the pre-test or post-test, but the treatment allocated to Group +M+P had a strong effect on their lexical variety throughout the treatment (see Figure 4.6 below).

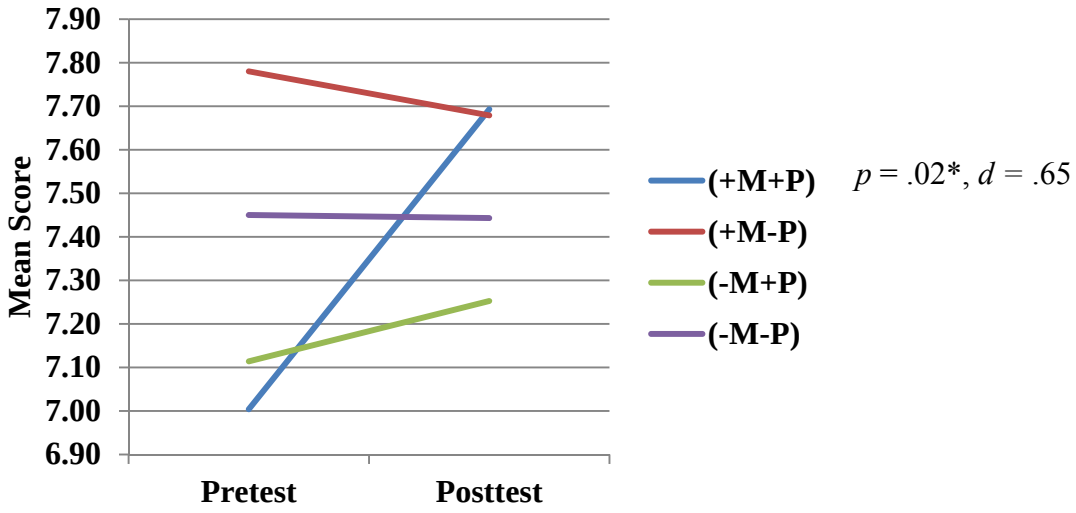


Figure 4.6 Complexity (Lexical variety)

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

On the whole, both syntactic and lexical variety increased significantly from the pre-test to the post-test for Group +M+P. That is, pre-task planners who had received metacognitive instruction made significant progress in producing a greater variety of verb forms and more lexical variety in their argumentative writing. In terms of Kellogg's (1996) model of writing, this may reflect the fact that these participants prioritised selecting the relevant lexical units needed to encode the propositional content of the text to the syntactic frames. It should be recalled that also in Levelt's (1989) model of speech production, lexis precedes and thus serves to prompt grammar and grammatical morphology during the formulation stage.

These findings are not in line with Ellis and Yuan's (2004) study in which pre-task planning when compared with no planning was found to have a strong effect on two complexity measures, i.e. syntactic complexity and syntactic variety, but no significant effect on lexical variety. If this had been the case, the pre-task planners in Group -M+P should have advantaged the same results, but they did not. Thus, it can be argued that the pre-post-test gains on syntactic and lexical variety was not only due to the provision of an opportunity for pre-task planning that was advantageous in terms of the syntactic and lexical variety of writing produced in Group +M+P, but it was also the combination of planning and the metacognitive strategy instruction that brought about significant progress during the treatment. This is compatible with the findings in the literature which suggest that the instructions learners are given can define the attention allocation between accuracy and complexity (e.g. Foster & Skehan, 1999). Similarly, the planning in this study was directed, through instructions, towards the content of the tasks rather than the language. This may explain why the trade-off is more evident between accuracy and complexity in this study.

Furthermore, whether planning instructions are guided or unguided and if guided whether learners have been directed to pre-plan either linguistic form, meaning, or both form and meaning, has revealed different results as to the extent to which learners seem to prioritise fluency, complexity or accuracy in their language production (Crookes, 1989; Ellis, 1987; Foster & Skehan, 1996, 1999; Mehnert, 1998; Wendel, 1997; Wigglesworth, 1997). It has been proposed that unguided planning in instructions leads to greater accuracy, whereas guided planning prioritises complexity in the language produced. In this study, the provision of guided planning was closely related to the instructional choices made during

the four sessions which could profitably channel participants' attention. That is, instructions helped direct participants' attention in Group +M+P to both the language, and the content and ideas to be expressed. Particularly the first session of the metacognitive instruction, which attempted to describe the different strategies that the participants could use during the planning phase of writing task was influential in guiding them to plan the meaning, vocabulary, and discourse structures of their writing. This meaning/form-focused strategic planning, when accompanied with other strategies of metacognitive training, was meant to help participants to develop their capacity for planning, monitoring and evaluating their writing. The combination of these strategies gave rise to participants' progress throughout the treatment with regard to greater syntactic variety and more lexical variety of language.

Summary

In this study, it was hypothesised that designing a metacognitive training program which focused on major metacognitive strategies required for writing, combined with effective planning instructions would lead L2 learners to produce better written performance. Quite contrary to the typical standard designs in task-based planning research, which require the participants in experimental and control groups to perform one single task under different planning conditions (e.g. strategic planning vs. within task planning vs. no planning), the design of this study attempted to address the acquisition of writing skills by the inclusion of a four-session treatment which was specifically aimed at identifying different changes in groups. In addition, their writing development was explored from a week prior to the treatment to a week afterwards. Overall, the findings in this study provided evidence for the benefits of metacognitive training, when combined with pre-task planning, on the enhancement of the syntactic and lexical variety of task-based performance on the one hand and the general writing performance of argumentative essays on the other hand.

CHAPTER 5. RESULTS AND DISCUSSION OF STUDY 2: RESEARCH QUESTIONS 1-3

This chapter aims to report and discuss the results of the quantitative analysis in Study 2. As in Study 1, the written data obtained from the pre-test, immediate and delayed post-tests were analysed both holistically and analytically with the aim of answering the following research questions:

1. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *general writing proficiency* of L2 learners' written products?
2. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *accuracy* of L2 learners' written products?
3. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *complexity* of L2 learners' written products?

5.1. Results

Before any analysis was conducted, the normal distribution of groups' scores on all variables was tested. Where normal distribution was established through a Kolmogorov-Smirnov test, a series of ANOVA was conducted, followed by post-hoc Tukey tests to explore differences between groups on the pre-test and (immediate/delayed) post-test performances. Where a normal distribution was not found, a Kruskal-Wallis test was used, followed by a Mann-Whitney U test as post-hoc analysis. The alpha for achieving statistical significance was set at .05. The results which did not show statistical significance are provided in Appendix F.

Before answering the research questions in detail, the results of the recruitment test used in the study are reported here. In order to make sure the participants were all homogenous, all participants' written texts were rated by two independent raters, a native speaker of

English and me. Overall, the mean scores of Rater One equalled 4.86 (*SD* = .32) and that of Rater Two was 4.87 (*SD* = .33).

Because the distribution of the raters' scores was significantly non-normal, Spearman's correlation was subsequently calculated to establish the degree of correlation between the two sets of writing scores. The Spearman's correlation coefficient obtained was .77 which was statistically significant (r = .77, p (2-tailed) = .00), suggesting a high degree of reliability. Accordingly, the scoring of the first rater was significantly correlated with that of the second rater and thus the inter-rater reliability was established on the dataset.

Scores were then entered into Kruskal-Wallis Test analysis with the alpha set at .05. The results revealed no significant differences between groups, whether the data were scored by the first rater (H(4) = 8.93, p = .06) or second rater (H(4) = 2.11, p = .71). Therefore, the Kruskal-Wallis Test confirmed that the five groups were homogenous with regard to their English writing proficiency and thus the random assignment of seventy participants to five groups was justified. A detailed report of the results of the holistic and detailed analyses corresponding to the three research questions is presented in Sections 5.1.1, 5.1.2, and 5.1.3.

5.1.1. Research Question 1

Effects of metacognitive strategy instruction and pre-task planning on the general writing proficiency of L2 learners' written production

The writing scores of the pre-test, immediate and delayed post-tests were initially compared based on their descriptive results. According to the descriptive statistics displayed in Table 5.1, all groups were almost equally clustered in the scores they gained in the pre-test and immediate post-test. With regard to the delayed post-test, Groups +M+P and +M-P gained the highest mean scores (M = 6.14, SD = .49 and M = 6.19, SD = .48 respectively) and appeared to perform better than the control group, with M = 5.56, SD = .53.

Table 5.1 Descriptive statistics: General writing proficiency

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	14	5.21	.32	.08
	+M-P	13	5.19	.48	.13
	-M+P	14	5.18	.31	.08
	-M-P	14	5.21	.25	.06
	Control	15	5.23	.37	.10
	Total	70	5.20	.34	.04
Immediate Post-test	+M+P	14	5.68	.46	.12
	+M-P	13	5.42	.40	.11
	-M+P	14	5.53	.53	.14
	-M-P	14	5.32	.37	.10
	Control	15	5.33	.40	.10
	Total	70	5.45	.44	.05
Delayed Post-test	+M+P	14	6.14	.49	.13
	+M-P	13	6.19	.48	.13
	-M+P	14	5.71	.42	.11
	-M-P	14	5.64	.41	.11
	Control	15	5.56	.53	.13
	Total	70	5.84	.52	.06

Note. M: Metacognitive instruction, P: Planning

Because the scores were non-normally distributed, the Kruskal-Wallis test, the non-parametric counterpart of one-way independent ANOVA, was used in order to specify whether any of the groups outperformed the other. Results revealed that none of the groups differed with regard to their performance in the pre-test (H (4) = .14, p = 1.00) or immediate post-test (H (4) = 5.85, p = .21). Differences between groups in the delayed post-test, however, were significant (H (4) = 16.84, p = .00).

As post-hoc tests, Mann-Whitney tests were used to follow up this significant finding in the delayed post-test (see Table 5.2). Altogether, ten comparisons were made among groups. A Bonferroni correction was applied and so all effects would be reported at a .05/10 = .005 critical level of significance. Only one significant difference was found: Group +M-P outperformed the control group (p = .005, d = .52). Furthermore, Group +M-P approached a significant tendency to outperform Group -M-P with a large effect size (p = .006, d = .53). Also, Group +M+P approached a significant tendency to

outperform the control group with an effect size almost at the threshold of Cohen's benchmark of large effect size (p = .008, d = .49).

Table 5.2 Mann-Whitney Tests: General writing proficiency in the delayed post-test

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)	Exact Sig. [2*(1-tailed Sig.)]
+M+P vs. +M-P	85.50	190.50	-.28	.778	.80
+M+P vs. -M+P	51.00	156.00	-2.27	.023	.03
+M+P vs. -M-P	45.00	150.00	-2.55	.011	.01
+M+P vs. Control group	46.50	166.50	-2.63	.008	.009
+M-P vs. -M+P	42.00	147.0	-2.49	.013	.017
+M-P vs. -M-P	36.50	141.5	-2.77	.006	.007
+M-P vs. Control group	39.00	159.00	-2.78	.005*	.006
-M+P vs. -M-P	91.00	196.0	-.35	.724	.769
-M+P vs. Control group	85.00	205.00	-.92	.355	.400
-M-P vs. Control group	92.50	212.50	-.57	.565	.565

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .005 level

In addition to the comparisons made between groups, the degree of increase or decrease from the pre-test to the post-tests was also measured. Friedman's ANOVA, the non-parametric alternative to the one-way repeated measures analysis of variance, was calculated. As shown in Table 5.3, the scores of all groups significantly changed over the course of time.

Table 5.3 Friedman Test: Pre-test to post-test progress in general writing proficiency

	+M+P	+M-P	-M+P	-M-P	Control group
N	14	13	13	14	15
Chi-Square	25.53	24.13	18.16	16.17	12.48
df	2	2	2	2	2
Asymp. Sig.	.00*	.00*	.00*	.00*	.00*

*. The significance is at the .05 level

Because the overall results from Friedman's ANOVA were significant (p < .05), Wilcoxon signed-rank tests were conducted as non-parametric post-hoc procedures. A Bonferroni correction was applied and so all effects would be reported at a .05/3 = .016 level of

significance. Tables 5.4 and 5.5 show the Wilcoxon signed-rank test statistics obtained from doing three comparisons, that is to say (1) the performance at the pre-test compared to the immediate post-test, (2) the performance at the pre-test compared to the delayed post-test, and (3) the performance at the immediate post-test compared to the delayed post-test. These results are reported in terms of test statistic T which is the smaller of the two sums of ranks for each test and the effect size *d*.

With regard to the experimental groups who received metacognitive training (see Table 5.4), it appeared that there were significant differences in performances. For Group +M+P, these differences were between the pre-test and the immediate post-test (T = 66, *d* = .59), between the pre-test and the delayed post-test (T = 105, *d* = .63), and between the immediate post-test and the delayed post-test (T = 78, *d* = .63). Similarly, Group +M-P progressed significantly from the pre-test to the immediate post-test (T = 21, *d* = .51), from the pre-test to the delayed post-test (T = 91, *d* = .70), and from the immediate post-test to the delayed post-test (T = 91, *d* = .64).

Table 5.4 Wilcoxon signed-rank tests: Groups +M+P and +M-P

	+M+P			+M-P		
	2 - 1	3 - 1	3 - 2	2 - 1	3 - 1	3 - 2
Z	-3.12	-3.37	-3.35	-2.45	-3.60	-3.27
Asymp. Sig. (2-tailed)	.002*	.001*	.001*	.014*	.000*	.001*

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .016 level

As shown in Table 5.5 below, the general writing performance of Group -M+P improved significantly from the pre-test to the immediate post-test (T = 28, *d* = .47), from the pre-test to the delayed post-test (T = 78, *d* = .63), but not from the immediate post-test to the delayed post-test (T = 15, medium-sized effect of *d* = .42). For Group -M-P, the outcome was somewhat different. This group did not significantly change from the pre-test to the immediate post-test (T = 12, small effect size of *d* = .25). Their performance, however, significantly improved from the pre-test to the delayed post-test (T = 56, *d* = .70), and from the immediate post-test to the delayed post-test (T = 45, *d* = .56).

Table 5.5 Wilcoxon signed-rank tests: Groups -M+P and -M-P and control group

	-M+P			-M-P			Control group		
	2 - 1	3 - 1	3 - 2	2 - 1	3 - 1	3 - 2	2-1	3 - 1	3 - 2
Z	-2.53	-3.35	-2.23	-1.34	-2.97	-3.00	-1.34	-2.64	-2.64
Asymp. Sig. (2-tailed)	.011	.001	.025	.180	.003	.003	.180	.008	.008

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .016 level

Finally, the Wilcoxon signed-rank tests for the control group, as demonstrated in Table 5.5 above, showed that they obtained similar results to Group -M-P. There were no significant changes from the pre-test to the immediate post-test (T = 12, small effect size of d = .24), but the performance significantly improved from the pre-test to the delayed post-test (T = 36, medium-sized effect of d = .48), and from the immediate post-test to the delayed post-test (T = 28, medium-sized effect of d = .48).

Summary

In terms of general writing proficiency, the between-group comparisons showed that overall only in the delayed post-test, the texts written by Group +M-P were significantly higher in quality than Group -M-P and the control group. Group +M+P also outscored the control group. As for pre-to-post-test analysis, all groups made significant progress.

5.1.2. Research Question 2

Effects of metacognitive strategy instruction and pre-task planning on the accuracy of L2 learners' written production

The focus in this section is on the results of investigating the second research question. In what follows the findings of the two different measures pertaining to analysing accuracy, i.e. error-free clauses and error-free verb forms, are presented.

5.1.2.1. Accuracy: Error-free clauses

Descriptive results for error-free clauses (syntactic, morphological, or lexical errors) for each group are presented in Table 5.6. All groups were closely clustered on the rate of error-free clauses. Scores ranged from 38% error free clauses (Group +M-P) to 44% (Groups +M+P and -M+P) in the pre-test; 39% (Group +M-P) to 47% (Group +M+P) in

the immediate post-test; and 43% (control group) to 51% (Group +M+P) in the delayed post-test. ANOVA results, however, showed that there were no significant differences between groups in terms of the accuracy of clauses in written productions in either pre-test, immediate post-test or delayed post-test ($p > .05$).

Performance on post-tests by all groups in terms of error-free clauses reflected a modest increase over time from the pre-test to each post-tests, with the exception of -M+P on the immediate post-test. It is interesting to note that Group +M+P made the greatest gains over all in terms of error-free clauses, in such a way that in the delayed post-test, half of all clauses were error-free. Also, the other group which received the metacognitive strategy training, i.e. Group +M-P, was the lowest at the outset, but increased by 6% to outperform the control group in the delayed post-test.

Table 5.6 Descriptive statistics: Error-free clauses

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	14	.44	.11	.03
	+M-P	13	.38	.08	.02
	-M+P	14	.44	.16	.04
	-M-P	14	.41	.07	.02
	Control	15	.40	.10	.02
	Total	70	.41	.11	.01
Immediate Post-test	+M+P	14	.47	.15	.04
	+M-P	13	.39	.10	.02
	-M+P	14	.43	.14	.03
	-M-P	14	.45	.09	.02
	Control	15	.42	.13	.03
	Total	70	.43	.12	.01
Delayed Post-test	+M+P	14	.51	.07	.02
	+M-P	13	.44	.09	.02
	-M+P	14	.47	.11	.03
	-M-P	14	.46	.10	.02
	Control	15	.43	.12	.03
	Total	70	.46	.10	.01

Note. M: Metacognitive instruction, P: Planning

Although no significant differences were detected between groups, repeated-measures ANOVA was calculated to ascertain whether the changes exhibited by each group from

the pre-test to the immediate post-test to the delayed post-test were statistically significant. In repeated-measures ANOVA, the effect of the experiment is shown up in the within-participant variance. A Bonferroni correction was also selected as a robust technique, in terms of power and control of the Type I error rate. The Mauchly's test (see Table 5.7) indicated that the assumption of sphericity had been met, $\chi 2(2) = .72$, $p = .70$.

Table 5.7 Mauchly's test of sphericity: Error-free clauses

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon: Greenhouse-Geisser	Epsilon: Huynh-Feldt	Epsilon: Lower-bound
	.99	.72	2	.70	.97	1.00	.50

Results of within-subjects effects revealed significant differences across the pre-test to the post-tests in some of the groups ($F(2, 138) = 15.07$, $p = .00$). The post-hoc analysis shown in Table 5.8 indicated that Group +M+P progressed significantly from the pre-test to the delayed post-test ($p = .01$, $d = .69$). Also, Group +M-P showed a tendency of significant progress in the delayed post-test compared to the pre-test ($p = .053$, $d = .62$).

Table 5.8 Groups +M+P and +M-P pairwise comparisons: Error-free clauses

	(I) Error-free Clauses	(J) Error-free Clauses	Mean Difference (I-J)	Std. Error	Sig.
+M+P	Pre-test	Immediate Post-test	-.03	.02	.58
		Delayed Post-test	-.07*	.02	.01
	Immediate Post-test	Pre-test	.03	.02	.58
		Delayed Post-test	-.03	.03	.68
	Delayed Post-test	Pre-test	.07*	.02	.01
		Immediate Post-test	.03	.03	.68
+M-P	Pre-test	Immediate Post-test	-.01	.01	1.00
		Delayed Post-test	-.06	.02	.053
	Immediate Post-test	Pre-test	.01	.01	1.00
		Delayed Post-test	-.04	.02	.12
	Delayed Post-test	Pre-test	.06	.02	.053
		Immediate Post-test	.04	.02	.12

Note. M: Metacognitive instruction, P: Planning

Post-hoc tests for the two non-metacognitive groups (with and without planning) and the control group showed no significant changes from the pre-test to the immediate post-test to the delayed post-test.

5.1.2.2. Accuracy: Error-free verb forms

Descriptive results for accurately used verb forms in terms of tense, aspect, modality, and subject-verb agreement in the pre-test and in the immediate and delayed post-tests are presented in Table 5.9 below. In the pre-test, the means of the accuracy of verb forms ranged from 68% in Group +M+P to 58% in Group +M-P. As with the immediate post-test, all groups seemed to have made slightly fewer errors, except for the control group which produced 62% error-free verbs compared to a 60% in the pre-test. In the delayed post-test, all groups appeared to have slightly increased, with the highest of 71% in Group +M+P and the lowest of 63% in the control group. The ANOVA results, however, showed that none of these small differences between groups in terms of the accuracy of verb forms in participants' written texts were significant in either pre-test or the two post-tests ($p > .05$).

Table 5.9 Descriptive statistics: Error-free verb forms

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	14	.68	.12	.03
	+M-P	13	.58	.08	.02
	-M+P	14	.64	.16	.04
	-M-P	14	.64	.10	.02
	Control	15	.60	.10	.02
	Total	70	.63	.12	.01
Immediate Post-test	+M+P	14	.67	.15	.04
	+M-P	13	.59	.10	.02
	-M+P	14	.63	.14	.04
	-M-P	14	.65	.09	.02
	Control	15	.62	.13	.03
	Total	70	.63	. 12	.01
Delayed Post-test	+M+P	14	.71	.07	.02
	+M-P	13	.64	.09	.02
	-M+P	14	.67	.11	.03
	-M-P	14	.69	.12	.03
	Control	15	.63	.12	.03
	Total	70	.67	.11	.01

Note. M: Metacognitive instruction, P: Planning

In order to determine the changes over time by each group, repeated-measures ANOVA was calculated. As shown in Table 5.10, the Mauchly's test indicated that the assumption of sphericity had been met, $\chi 2(2) = 2.23$, $p = .32$.

Table 5.10 Mauchly's test of sphericity: Error-free verb forms

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon Greenhouse-Geisser	Epsilon Huynh-Feldt	Epsilon Lower-bound
	.96	2.23	2	.32	.97	.99	.50

According to the results of within-subjects effects, there were no significant differences across the pre-test to the post-tests in any of the groups $F(2) = 11.13$, $p = .09$. Thus, no post-hoc pairwise comparisons were conducted.

Summary

As for accuracy, the ANOVA results failed to show the between-group differences to be statistically significant in the case of both variables of error-free clauses and error-free verb forms. The within-group comparisons over time did not return significant results for the two measures either, except for Group +M+P who progressed in the provision of error-free clauses in the delayed post-test, compared to the pre-test.

5.1.3. Research Question 3

Effects of metacognitive strategy instruction and pre-task planning on the complexity of L2 learners' written production

In this section, the results of the three measures pertaining to analysing complexity, i.e. syntactic complexity, syntactic variety, and lexical variety, are presented.

5.1.3.1. Complexity: Syntactic complexity

The descriptive results presented in Table 5.11 show that, on average, the five groups were closely clustered when the syntactic complexity of their written production were measured by the ratio of clauses to T-units in the pre-test. Results of the one-way ANOVA analysis confirmed that the subtle differences between groups were not significant in the syntactic complexity of participants' production in any of the three tests ($p > .05$).

With regard to the immediate post-test, all groups appeared to be performing very similarly to the pre-test. However, the delayed post-test results showed a somewhat different picture. All groups appeared to have increased, however the control group tended to slightly decrease from the immediate post-test (*M* = 2.01, *SD* = .49) to the delayed post-test (*M* = 1.94, *SD* = .36).

Table 5.11 Descriptive statistics: Syntactic complexity

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	14	1.96	.42	.11
	+M-P	13	1.97	.37	.10
	-M+P	14	1.90	.41	.11
	-M-P	14	1.99	.42	.11
	Control	15	1.89	.26	.06
	Total	70	1.94	.37	.04
Immediate Post-test	+M+P	14	1.88	.26	.06
	+M-P	13	1.94	.28	.07
	-M+P	14	2.02	.49	.13
	-M-P	14	2.02	.46	.12
	Control	15	2.01	.49	.12
	Total	70	1.98	.40	.04
Delayed Post-test	+M+P	14	2.20	.47	.12
	+M-P	13	2.31	.48	.13
	-M+P	14	2.23	.51	.13
	-M-P	14	2.28	.49	.13
	Control	15	1.94	.36	.09
	Total	70	2.19	.47	.05

Note. M: Metacognitive instruction, P: Planning

In order to determine the changes over time by each individual group, repeated-measures ANOVA was calculated. As shown in Table 5.12, the Mauchly's test indicated that the assumption of sphericity was violated, $\chi 2(2) = 9.30$, $p = .01$. Thus it was reasonable to conclude that the variances of differences between the pre-test, the immediate and the delayed post-tests were significantly different (i.e. they were not roughly equal). SPSS analysis produced two corrections based upon the estimates of sphericity advocated by Greenhouse and Geisser (1959) and Huynh and Feldt (1976). Both of these estimates gave rise to a correction factor that was applied to the degrees of freedom used to assess the

observed F-ratio. According to Field (2009), the closer the Greenhouse–Geisser correction ε^ is to 1, the more homogeneous the variances of differences, and thus the closer the data are to being spherical. In Study 2, there were three conditions of pre-test, immediate post-test and delayed post-test, therefore the lower limit of ε^ was 1/(3 − 1), or .5 which is the lower-bound in Table 5.12. The results showed that the calculated value of ε^ was .88. This was closer to the upper limit of 1 than the lower limit of .5 and it therefore represented only a slight deviation from sphericity.

Table 5.12 Mauchly's test of sphericity: Syntactic complexity

					Epsilon		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
	.87	9.30	2	.01	.88	.90	.50

Having corrected degrees of freedom using Greenhouse and Geisser estimates of sphericity (ε = .88), the results of within-subjects effects then showed that there were significant differences from the pre-test to the two post-tests, F(1.87, 122.35) = 18.72, p = .00. Table 5.13 shows the post-hoc tests for the repeated-measures syntactic complexity variable for the metacognitive groups (with and without planning). By looking at the mean scores and the significance values, the performance of Group +M+P reflected greater syntactic complexity in the delayed post-test compared to the pre-test (p = .02, d = .66). Group +M-P demonstrated increased complexity both from the pre-test to the delayed post-test (p = .00, d = .81), and from the immediate post-test to the delayed post-test (p = .00, d = .73), but not from the pre-test to the immediate post-test (p = 1.00, small d = .11).

Table 5.13 Groups +M+P and +M-P pairwise comparisons: Syntactic complexity

	(I) Syntactic Complexity	(J) Syntactic Complexity	Mean Difference (I-J)	Std. Error	Sig.
+M+P	Pre-test	Immediate Post-test	.08	.13	1.00
		Delayed Post-test	-.23*	.07	.02
	Immediate Post-test	Pre-test	-.08	.13	1.00
		Delayed Post-test	-.31	.14	.14
	Delayed Post-test	Pre-test	.23*	.07	.02
		Immediate Post-test	.31	.14	.14
+M-P	Pre-test	Immediate Post-test	.02	.05	1.00
		Delayed Post-test	-.34*	.07	.00
	Immediate Post-test	Pre-test	-.02	.05	1.00
		Delayed Post-test	-.36*	.09	.00
	Delayed Post-test	Pre-test	.34*	.07	.00
		Immediate Post-test	.36*	.09	.00

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

The post-hoc tests for the non-metacognitive groups (with and without planning), presented in Table 5.14, show that Group –M+P improved in all comparisons related to time: from the pre-test to the immediate post-test (p = .04, d = .61), from the pre-test to the delayed post-test (p = .00, d = .87), and from the immediate post-test to the delayed post (p = .00, d = .80). The performance of Group -M-P improved in two comparisons: from the pre-test to the delayed post-test (p = .00, d = .74), and from the immediate post-test to the delayed post-test (p = .00, d = .75), but not from the pre-test to the immediate post-test. The pairwise comparisons for the control group, however, uncovered no significant differences between the pre-test, immediate post-test, and delayed post-test.

Table 5.14 Groups -M+P, -M-P, and control group pairwise comparisons: Syntactic complexity

	(I) Syntactic Complexity	(J) Syntactic Complexity	Mean Difference (I-J)	Std. Error	Sig.
-M+P	Pre-test	Immediate Post-test	-.12*	.04	.04
		Delayed Post-test	-.33*	.05	.00
	Immediate Post-test	Pre-test	.12*	.04	.04
		Delayed Post-test	-.21*	.04	.00
	Delayed Post-test	Pre-test	.33*	.05	.00
		Immediate Post-test	.21*	.04	.00
-M-P	Pre-test	Immediate Post-test	-.03	.08	1.00
		Delayed Post-test	-.29*	.07	.00
	Immediate Post-test	Pre-test	.03	.08	1.00
		Delayed Post-test	-.26*	.06	.00
	Delayed Post-test	Pre-test	.29*	.07	.00
		Immediate Post-test	.26*	.06	.00
Control group	Pre-test	Immediate Post-test	-.12	.12	1.00
		Delayed Post-test	-.04	.10	1.00
	Immediate Post-test	Pre-test	.12	.12	1.00
		Delayed Post-test	.07	.13	1.00
	Delayed Post-test	Pre-test	.04	.10	1.00
		Immediate Post-test	-.07	.13	1.00

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

5.1.3.2. Complexity: Syntactic variety

Table 5.15 shows the descriptive results of each group's total number of different grammatical verb forms used in the pre-test, immediate and delayed post-tests. By looking at the mean scores of groups, we can see that all groups appeared to have been closely clustered in the three tests. This was confirmed by the results of ANOVA which showed that there were no significant differences among groups in terms of syntactic variety of production in either pre-test, or the two post-test writing ($p > .05$).

Table 5.15 Descriptive statistics: Syntactic variety

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	14	7.86	1.83	.49
	+M-P	13	7.92	1.11	.31
	-M+P	14	6.71	1.32	.35
	-M-P	14	7.29	1.43	.38
	Control	15	7.20	1.42	.36
	Total	70	7.39	1.47	.17
Immediate Post-test	+M+P	14	8.64	1.39	.37
	+M-P	13	8.77	1.09	.30
	-M+P	14	7.79	1.12	.30
	-M-P	14	7.93	1.59	.42
	Control	15	7.60	1.64	.42
	Total	70	8.13	1.43	.17
Delayed Post-test	+M+P	14	8.71	1.49	.39
	+M-P	13	9.08	1.49	.41
	-M+P	14	7.86	1.65	.44
	-M-P	14	8.29	1.54	.41
	Control	15	7.87	1.84	.47
	Total	70	8.34	1.64	.19

Note. M: Metacognitive instruction, P: Planning

In order to determine the changes over time by each individual group, repeated-measures ANOVA was calculated. The Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2(2) = 7.82$, $p = .02$ (see Table 5.16). However, Greenhouse and Geisser's correction applied to the degrees of freedom showed only a slight deviation from the sphericity ($\varepsilon\hat{} = .90$).

Table 5.16 Mauchly's test of sphericity: Syntactic variety

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon: Greenhouse-Geisser	Epsilon: Huynh-Feldt	Epsilon: Lower-bound
	.89	7.82	2	.02	.90	.92	.50

According to the results of within-subjects effects, there were significant differences from the pre-test to the post-tests, $F(1.80, 124.46) = 24.70$, $p = .00$. In order to ascertain the location of significance, the post-hoc tests were calculated for the repeated-measures variable. Table 5.17 shows the pairwise comparisons for the metacognitive groups (with

and without planning). The mean differences and the significance values demonstrate that Group +M+P progressed significantly in one pairwise comparison: from the pre-test to the delayed post-test (p = .05, d = .60). Group +M-P produced syntactically more varied texts in both immediate post-test (p = .00, d = .78) and delayed post-test (p = .02, d = .66) compared to the pre-test.

Table 5.17 +M+P and +M-P pairwise comparisons: Syntactic variety

	(I) Syntactic Variety	(J) Syntactic Variety	Mean Difference (I-J)	Std. Error	Sig.
+M+P	Pre-test	Immediate Post-test	-.78	.55	.54
		Delayed Post-test	-.85*	.31	.05
	Immediate Post-test	Pre-test	.78	.55	.54
		Delayed Post-test	-.07	.49	1.00
	Delayed Post-test	Pre-test	.85*	.31	.05
		Immediate Post-test	.07	.49	1.00
+M-P	Pre-test	Immediate Post-test	-.84*	.19	.00
		Delayed Post-test	-1.15*	.37	.02
	Immediate Post-test	Pre-test	.84*	.19	.00
		Delayed Post-test	-.30	.38	1.00
	Delayed Post-test	Pre-test	1.15*	.37	.02
		Immediate Post-test	.30	.38	1.00

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

The post-hoc tests pairwise comparisons for the non-metacognitive groups (with and without planning) and the control group, presented in Table 5.18, show that the performance of both non-metacognitive groups increased not only from the pre-test to the immediate post-test (p = .00, d = .80 for Group -M+P, and p = .04, d = .62 for Group -M-P), but also from the pre-test to the delayed post-test (p = .00, d = .87 for Group -M+P, and p = .00, d = .73 for Group -M-P). However, the control group revealed no significant differences in pairwise comparisons from the pre-test to the immediate post-test to the delayed post-test.

Table 5.18 -M+P, -M-P and control group pairwise comparisons: Syntactic variety

	(I) Syntactic Variety	(J) Syntactic Variety	Mean Difference (I-J)	Std. Error	Sig.
-M+P	Pre-test	Immediate Post-test	-1.07*	.22	.00
		Delayed Post-test	-1.14*	.17	.00
	Immediate Post-test	Pre-test	1.07*	.22	.00
		Delayed Post-test	-.07	.32	1.00
	Delayed Post-test	Pre-test	1.14*	.17	.00
		Immediate Post-test	.07	.32	1.00
-M-P	Pre-test	Immediate Post-test	-.64*	.22	.04
		Delayed Post-test	-1.00*	.25	.00
	Immediate Post-test	Pre-test	.64*	.22	.04
		Delayed Post-test	-.35	.29	.71
	Delayed Post-test	Pre-test	1.00*	.25	.00
		Immediate Post-test	.35	.29	.71
Control group	Pre-test	Immediate Post-test	-.40	.19	.16
		Delayed Post-test	-.66	.27	.08
	Immediate Post-test	Pre-test	.40	.19	.16
		Delayed Post-test	-.26	.35	1.00
	Delayed Post-test	Pre-test	.66	.27	.08
		Immediate Post-test	.26	.35	1.00

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

5.1.3.3. Complexity: Lexical variety

Lexical variety was measured by Giraud index of lexical richness. The descriptive results, presented in Table 5.19, show that Group +M+P produced the most lexically varied writing in the pre-test (M = 7.79, SD = .71), and the control group the least (M = 7.11, SD = .64). Groups +M-P, -M+P, and -M-P appeared to have been closely clustered. In the immediate post-test, Groups +M+P, +M-P, and -M+P tended to increase, however Group -M-P showed a decrease (from M = 7.42, SD = .77 in the pre-test to M = 7.20, SD = .79 in the immediate post-test), and the performance of the control group remained almost unchanged (M = 7.11, SD = .64, and M = 7.12, SD = .67 in the pre-test and the immediate post-test respectively). With regard to the delayed post-test, the performance of all groups slightly increased in terms of lexical variety, particularly Group +M+P (M = 8.28, SD = .57) and the control group (M = 7.63, SD = .84). However, the degree of variety in lexical choices in Groups +M-P and -M+P slightly decreased, compared to the immediate post-test (M = 7.72, SD = .80, and M = 7.59, SD = .61 respectively).

Table 5.19 Descriptive statistics: Lexical variety

		N	Mean	Std. Deviation	Std. Error
Pre-test	+M+P	14	7.79	.71	.18
	+M-P	13	7.40	.81	.22
	-M+P	14	7.38	.60	.16
	-M-P	14	7.42	.77	.20
	Control	15	7.11	.64	.16
	Total	70	7.41	.72	.08
Immediate Post-test	+M+P	14	8.22	.71	.19
	+M-P	13	7.85	.70	.19
	-M+P	14	7.69	.65	.17
	-M-P	14	7.20	.79	.21
	Control	15	7.09	.67	.17
	Total	70	7.60	.80	.09
Delayed Post-test	+M+P	14	8.28	.57	.15
	+M-P	13	7.72	.80	.22
	-M+P	14	7.59	.61	.16
	-M-P	14	7.53	.70	.19
	Control	15	7.63	.84	.21
	Total	70	7.75	.74	.09

Note. M: Metacognitive instruction, P: Planning

As presented in Table 5.20, there were no significant differences among groups in the pre-test. However, there was a significant difference among groups in the immediate post-test (p = .00, d = .52). There also appeared to exist a near-significant difference between groups in the delayed post-test (p = .051, with a medium effect size of d = .36).

Table 5.20 ANOVA: Lexical variety

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	3.39	4	.85	1.67	.16
	Within Groups	32.87	65	.50		
	Total	36.27	69			
Immediate Post-test	Between Groups	12.42	4	3.10	6.19	.00
	Within Groups	32.60	65	.50		
	Total	45.03	69			
Delayed Post-test	Between Groups	5.15	4	1.29	2.50	.051
	Within Groups	33.48	65	.51		
	Total	38.63	69			

*. The mean difference is significant at the .05 level.

In order to specify where the differences lay in the immediate post-test, post-hoc Tukey test analysis (Table 5.21) indicated that Group +M+P produced more lexically varied writing than Group -M-P (p = .00, d = .57) and the control group (p = .00, d = .64). Group +M-P also outperformed the control group (p = .05, d = .49). In the delayed post-test, again Group +M+P showed a strong tendency to outperform Group -M-P (p = .056, d = .51). In all these three significant between-group differences, the effect sizes were large.

Table 5.21 Tukey: Lexical variety (Immediate post-test)

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
+M+P	+M-P	.37	.27	.64
	-M+P	.53	.26	.27
	-M-P	1.02*	.26	.00
	Control	1.12*	.26	.00
+M-P	+M+P	-.37	.27	.64
	-M+P	.16	.27	.97
	-M-P	.65	.27	.13
	Control	.75*	.26	.05
-M+P	+M+P	-.53	.26	.27
	+M-P	-.16	.27	.97
	-M-P	.48	.26	.37
	Control	.59	.26	.17
-M-P	+M+P	-1.02*	.26	.00
	+M-P	-.65	.27	.13
	-M+P	-.48	.26	.37
	Control	.10	.26	.99
Control group	+M+P	-1.12*	.26	.00
	+M-P	-.75*	.26	.05
	-M+P	-.59	.26	.17
	-M-P	-.10	.26	.99

Note. M: Metacognitive instruction, P: Planning

*. The mean difference is significant at the .05 level.

In order to determine the changes over time by each individual group, repeated-measures ANOVA was calculated. The Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2(2) = 13.87$, p = .00 (see Table 5.22). The calculated value of ε^ is .84 and it therefore represents a slight deviation from sphericity.

Table 5.22 Mauchly's test of sphericity: Lexical variety

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
	.81	13.87	2	.00	.84	.86	.50

According to the results of within-subjects effects in Table 5.23, there were significant differences from the pre-test to the post-tests, $F(1.68, 116.50) = 10.22$, $p = .00$. By looking at the post-hoc tests shown in Table 5.23, we can see that Group +M+P progressed significantly from the pre-test to the immediate post-test ($p = .04$, $d = .62$) and from the pre-test to the delayed post-test ($p = .05$, $d = .60$). Also, Group +M-P progressed producing lexically varied writing in the immediate post-test compared to the pre-test ($p = .01$, $d = .72$).

Table 5.23 Groups +M+P and +M-P pairwise comparisons: Lexical variety

	(I) Lexical Variety	(J) Lexical Variety	Mean Difference (I-J)	Std. Error	Sig.
+M+P	Pre-test	Immediate Post-test	-.43*	.15	.04
		Delayed Post-test	-.48*	.17	.05
	Immediate Post-test	Pre-test	.43*	.15	.04
		Delayed Post-test	-.05	.11	1.00
	Delayed Post-test	Pre-test	.48*	.17	.05
		Immediate Post-test	.05	.11	1.00
+M-P	Pre-test	Immediate Post-test	-.45*	.12	.01
		Delayed Post-test	-.32	.27	.76
	Immediate Post-test	Pre-test	.45*	.12	.01
		Delayed Post-test	.12	.21	1.00
	Delayed Post-test	Pre-test	.32	.27	.76
		Immediate Post-test	-.12	.21	1.00

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

According to Table 5.24, the pairwise comparisons for the non-metacognitive groups (with and without planning) showed that only Group -M+P progressed from the pre-test to the immediate post-test ($p = .00$, $d = .65$), but the comparisons over time did not uncover any changes for Group -M-P. Surprisingly, the control group revealed a significant progress not only from the pre-test to the delayed post-test ($p = .04$, $d = .60$), but also from the immediate post-test to the delayed post-test ($p = .04$, $d = .60$).

Table 5.24 Groups -M+P, -M-P and control group pairwise comparisons: Lexical variety

	(I) Lexical Variety	(J) Lexical Variety	Mean Difference (I-J)	Std. Error	Sig.
-M+P	Pre-test	Immediate Post-test	-.30*	.09	.02
		Delayed Post-test	-.20	.13	.46
	Immediate Post-test	Pre-test	.30*	.09	.02
		Delayed Post-test	.10	.15	1.00
	Delayed Post-test	Pre-test	.20	.13	.46
		Immediate Post-test	-.10	.15	1.00
-M-P	Pre-test	Immediate Post-test	.22	.09	.10
		Delayed Post-test	-.10	.15	1.00
	Immediate Post-test	Pre-test	-.22	.09	.10
		Delayed Post-test	-.32	.14	.11
	Delayed Post-test	Pre-test	.10	.15	1.00
		Immediate Post-test	.32	.14	.11
Control group	Pre-test	Immediate Post-test	.01	.04	1.00
		Delayed Post-test	-.52*	.18	.04
	Immediate Post-test	Pre-test	-.01	.04	1.00
		Delayed Post-test	-.53*	.19	.04
	Delayed Post-test	Pre-test	.52*	.18	.04
		Immediate Post-test	.53*	.19	.04

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

Summary

Three variables were assessed to measure the complexity of language use in participants' written texts: syntactic complexity, syntactic variety, and lexical variety. No significant differences were found among groups in terms of the syntactic complexity and syntactic variety. In the case of the lexical variety, the two metacognitive groups outperformed the control group, while the metacognitive (with planning) group outscored the non-metacognitive (without planning) group. Interestingly, all groups including the control group advanced in their ability to vary lexical choices over time.

5.1.4. Summary of results

Tables 5.25 and 5.26 summarise the results obtained from the quantitative data analysis in Study 2. As demonstrated in these tables, no significant differences between groups were

noted in the pre-test and immediate post-test in any of the measures, with the exception of the production of lexically varied writing in the immediate post-test, where Group +M+P outperformed both Group -M-P and the control group. Also, Group +M-P outscored the control group. In the delayed post-test, all groups performed similarly with regard to the measures investigated in this study, except for general writing proficiency. With respect to this measure, Group +M+P outperformed the control group, and Group +M-P outperformed Group -M-P and the control group.

Table 5.25 Summary of between-group results in Study 2

Measures		Pre-test	Immediate Post-test	Delayed Post-test
General writing proficiency		n.s.	n.s.	+M+P > Control +M-P > -M-P +M-P > Control
Accuracy	Error-free clauses	n.s.	n.s.	n.s.
	Error-free verbs	n.s.	n.s.	n.s.
Complexity	Syntactic complexity	n.s.	n.s.	n.s.
	Syntactic variety	n.s.	n.s.	n.s.
	Lexical variety	n.s.	+M+P > -M-P +M+P > Control +M-P > Control	+M+P > -M-P

The results gained to show the progress from the pre-test to the two post-tests indicated that except for the error-free verb form measures, where none of the groups manifested any progress rate, other measures revealed significant changes over time in within-group comparisons.

Table 5.26 Summary of within-group results in Study 2

Measures		Progress (Pre-Posts)
General writing proficiency		+M+P (Delayed Post > Immediate Post > Pre-test) +M-P (Delayed Post > Immediate Post > Pre-test) -M+P (Delayed Post, Immediate Post > Pre-test) -M-P (Delayed Post > Pre-test, Immediate Post) Control (Delayed Post > Pre-test, Immediate Post)
Accuracy	Error-free clauses	+M+P (Delayed Post > Pre-test)
	Error-free verbs	n.s.
Complexity	Syntactic complexity	+M+P (Delayed Post > Pre-test) +M-P (Delayed Post > Pre-test, Immediate Post) -M+P (Delayed Post > Immediate Post > Pre-test) -M-P (Delayed Post > Pre-test, Immediate Post)
	Syntactic variety	+M+P (Delayed Post > Pre-test) +M-P (Delayed Post, Immediate Post > Pre-test) -M+P (Delayed Post, Immediate Post > Pre-test) -M-P (Delayed Post, Immediate Post > Pre-test)
	Lexical variety	+M+P (Delayed Post , Immediate Post > Pre-test) +M-P (Immediate Post > Pre-test) -M+P (Immediate Post > Pre-test) Control (Delayed Post > Pre-test, Immediate Post)

5.2. Discussion of results

Study 2 investigated the effects of pre-task planning and metacognitive strategy training on EFL learners' writing ability. Although there has been some research investigating the impact of planning in writing contexts, there is little on metacognitive instruction and no firm evidence to demonstrate that metacognitive strategy instruction, when combined with writing instruction, may promote L2 learners' written production. In this section, first the following research questions will be answered.

1. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *general writing proficiency* of L2 learners' written products?
2. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *accuracy* of L2 learners' written products?

3. In a written argumentative task, what are the effects of metacognitive strategy instruction and pre-task planning on the *complexity* of L2 learners' written products?

The answers to these research questions are followed by an in-depth discussion of the results concerning the different measures of writing. In accomplishing this, I connect the findings to previous literature on strategy training and task planning in L2 writing.

5.2.1. Research Question 1

Effects of metacognitive strategy instruction and pre-task planning on the general writing proficiency of L2 learners' written production

To address this research question, a holistic analysis of the written data was performed to determine the general writing ability of the participants at three different phases of time. The answer to this research question is split into two parts: between-group and within-group comparisons. The results of the holistic analysis, as shown in Figure 5.1, indicated that the comparison between groups did not achieve statistical significance in the pre-test, so it was ensured that all groups performed similarly. In the immediate post-test, no significant differences were detected among groups either. This shows that all groups demonstrated similar improvements in their writing. In the delayed post-test, however, the answer to the research question can be partially confirmed: overall, Group +M-P significantly outscored the control group. The more highly skilled performance of Group +M-P compared to Group -M-P, and the superior performance of Group +M+P compared to the control group also showed a tendency to significance. In other words, the metacognitive instruction was not more effective in the overall development of participants' subsequent writing performance which was completed immediately after the treatment, but it proved more effective in the long term, that is to say, in a new piece of writing four weeks after the treatment. Quite large effect sizes were noted in the delayed post for these consequential differences which indicate a significant and meaningful effect of the metacognitive instruction, albeit not pre-task planning, on the general writing proficiency. Thus, the answer to this part of the research question is affirmative, in that metacognitive instruction can explain the superior performance of metacognitively trained groups in the delayed post-test.

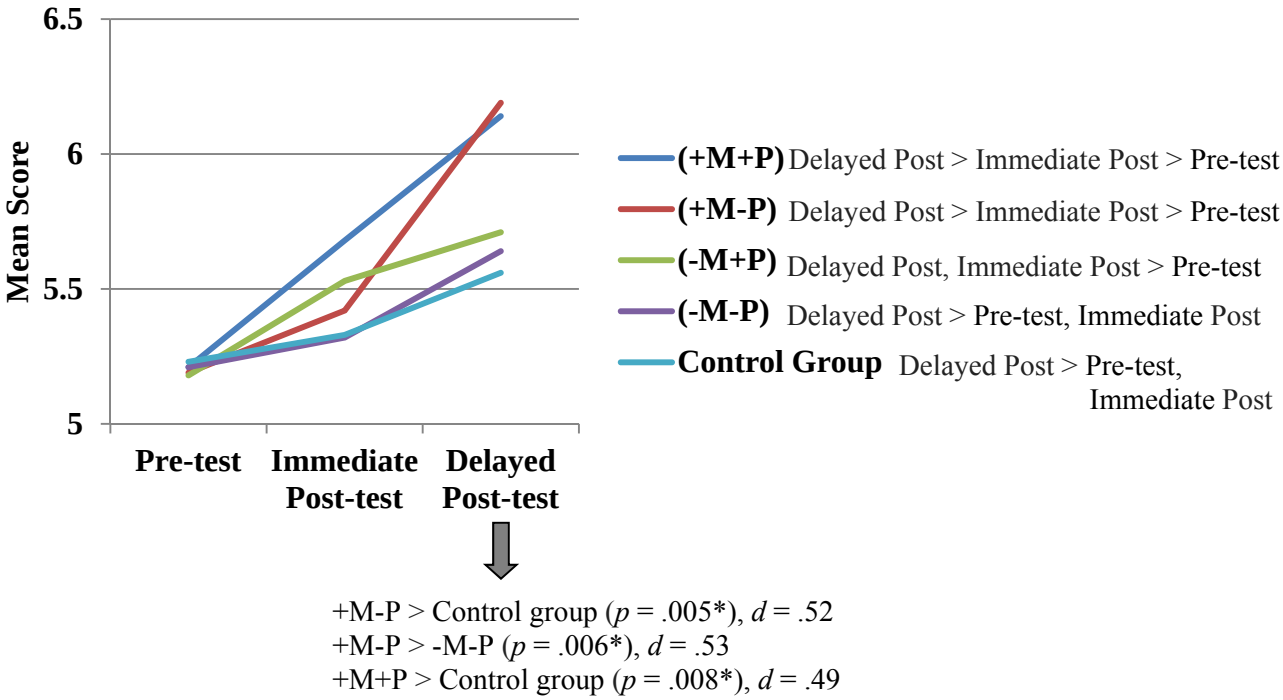


Figure 5.1 General writing proficiency

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .005 level

The results gained from within-group comparisons, as shown in Figure 5.1 above, indicated that the two metacognitively instructed groups showed significant progress in the three comparisons made over time, that is to say, from the pre-test to the immediate post-test, from the pre-test to the delayed post-test, and from the immediate post-test to the delayed post-test. One can, thus, conclude that the impact of treatment over time was a resounding success for these two experimental groups. Rather than showing that teaching metacognitive strategies was the sole cause of this improvement, the results suggest a more complex picture for Group -M+P which significantly progressed from the pre-test to the immediate post-test, and from the pre-test to the delayed post-test. Group -M-P and the control group did not significantly change from the pre-test to the immediate post-test. They significantly progressed, however, from the pre-test to the delayed post-test, and from the immediate post-test to the delayed post-test. The answer to this part of the research question is therefore not affirmative, because the progress over time cannot be attributed to either of the two independent variables.

Before discussing possible explanations for these results, it is necessary to note that since July 2007, scores for the IELTS Writing Module have been reported in whole or half-bands in the same way as for the Reading and Listening modules. The assessment criteria and the way examiners assess candidates' performances have not changed. What has changed is the reporting of the scores on a more precise scale. Since half-bands were introduced to allow for greater differentiation within a band, this change was introduced as a service to stakeholders who required more information on candidates' abilities. When a score for a particular skill is averaged, it is rounded up or down to the closest appropriate band scale as with scores for the four IELTS skill tests to arrive at an overall IELTS band score. Consequently, a 6.5 in Writing, for instance, represents a very strong performance at the Band 6 level. One may argue that the difference between 6 and 6.5 is measurable and valid, but the difference between 6 and 6.15 does not indicate that one candidate is better than the other. It might be argued that just because scores are significantly different does not mean that they truly indicate a difference or are meaningful in terms of describing the writing skill.

The current study, however, adheres to the experimental interpretation of results and the effect sizes they show, rather than being governed by the actual operationalisation of the IELTS band use, particularly since sample tests, rather than official IELTS tests, were used in this study. Even if intended to interpret within IELTS band scale, the two metacognitive groups' mean scores in the delayed post-test (*M* (Group +M+P) = 6.14 and *M* (Group +M-P) = 6.19) would be both practically reported with the score of 6 according to the IELTS scoring scale which were still significantly higher than the control group (*M* = 5.56). However, they would not actually have meaning as different scores, since they would be within the one band; in other words, they would still represent the same broad category of performance.

- ***Effects of metacognitive instruction on general writing proficiency***

One possibility for the superior performance of metacognitive groups in the delayed post-test is that, as Graham and Harris (2007) have asserted, it is not expected that all students will master strategies such as planning at the same time. Some students clearly take longer to learn the prerequisite knowledge and processes and to independently apply the strategies in further tasks. Another possibility is that the metacognitive instruction trained participants to gain conscious awareness of their micro- and macro-level knowledge of

writing, and to acquire the conscious ability to control and manipulate the related cognitive processes. As a result, they determinedly attended to certain rules regarding the use of a wide range of vocabulary resources and grammatical structures which they subsequently used, but needed time to be able to perform those rules effectively. That could explain why they obtained higher band scores for *Lexical Resource* and *Grammatical Range and Accuracy* in the IELTS writing scale (see Appendices D & E). In order to unpack these two components of this band scale, I note that the criterion of *Lexical Resource* refers to the range of vocabulary the candidate has used, and the accuracy and appropriacy of that use in terms of the specific task. It is also important to stress that the *Grammatical Range and Accuracy* criterion refers to the range and accurate use of the candidate's grammatical resource as manifested in the candidate's writing at sentence level. However, metacognition did more than just focus on grammatical errors: it also served to draw participants' attention to higher order stylistic and organisational errors. In addition to the two sentence-level criteria above, metacognition helped them to focus on two other macro-level bands of this writing scale: C*ohesion and Coherence* and *Task Response*. Coherence is concerned with the overall clarity and fluency of the message: how the response organises and links information, ideas and language through logical sequencing, whereas cohesion refers to the varied and appropriate use of cohesive devices (for example logical connectors, pronouns, and conjunctions) to assist in making the conceptual and referential relationships between and within sentences discernibly clear. The final criterion of *Task Response* assesses how fully the candidate addresses all parts of the task, and presents a fully developed position in answer to the question with relevant, fully extended and well-supported ideas. In short, the metacognitive strategy instruction showed to have made a significant impact on participants' micro- and macro-quality of writing over time.

In a study analysing the cognitive processing behaviours of anglophone French undergraduate students, Whalen and Menard (1995) showed that while writing an argumentative text in L2 (French), they attended to the lexical and morphosyntactic levels of the writing process at the expense of the textual and rhetorical components. This might explain why the participants of my study performed almost equally in the immediate post-test. It is possible that they were not yet fully proficient writers, and paid attention to the lexical and syntactic search to a greater extent than to the higher level components. At this stage, however, it is important to underline the fact that this claim needs to be applied

cautiously, and it will be unfolded in greater depth further on in relation to the micro-level skills. This will make the turn of my conclusion from this result more apparent.

These findings support those of previous studies such as Y. Zhang (2010) in which metacognitive knowledge was shown to have facilitating effects on English writing of Chinese EFL learners. Similarly, in an empirical study investigating metacognitive strategies-based instruction for Chinese vocational college students, Lv and Chen (2010) found that the experimental group outperformed the control group in their post-test writing performances. Overall, the primary findings in Graham's (2006) meta-analysis of studies on strategy instruction and teaching of writing illustrated that strategy instruction which taught one or more strategies for planning, revising, or editing text not only had a significant impact on students' writing immediately after the instruction, but also the effects were maintained over time and were generalised to new tasks and situations.

It appeared that almost all participants in the present study acquired sufficient knowledge about the nature of argumentative writing tasks over time and about the types of strategies required to accomplish the tasks through practice either by working on the task type and/or by classes external to the data collection. The overall progress of all groups over time, with metacognitively trained groups progressing in all three time comparisons, and the three other groups in only two time comparisons, may lead us to consider the role of other possible factors including maturation, history and exposure. This observation is strengthened by Gosden (1996) who argued, concerning the writing activities of Japanese novice researchers in his study, that the inevitable effects of maturation of L2 writing expertise were manifested in the development of linguistic and socio-rhetorical control over a new genre of writing. Similarly, in the present study it is tentatively speculated that participants' writing practice helped them to develop the abilities required when confronted with argumentative writing. The critical role of practice in the development of writing ability has also been studied by Manchón and Roca de Larios (2007) and Sasaki (2004). The former scholars found connection between writing proficiency, time allocated to planning, and skill in dealing with high-level aspects of text. Accordingly, the more proficient the learners were, the more time they allocated to planning and the better they dealt with high-level textual matters while in the process of planning. Sasaki's (2004) study also discovered that the composition scores of both ESL and EFL groups significantly increased after taking two semesters of process-writing instruction.

To reinforce the possibility of practice effect, it would be worthwhile to study the changing patterns over-time that the control group experienced. This group was in fact part of the whole educational institution in which the study was being carried out. This particular language school offered courses individually for the four required skills of the IELTS test, namely Listening, Speaking, Reading and Writing, without obliging students to enrol in more than one course. Therefore, the respective participants in the control group could have made progress independent of the study through involvement in writing classes, or alternatively could have been enrolled in one or more classes for different skills apart from writing. In short, it could be argued that although metacognition could stand out as an independently effective variable, it may promote learning along with other contributing factors such as maturation and practice effects.

- ***Effects of pre-task planning on general writing proficiency***

It was hypothesised that repeatedly reinforcing the use of pre-writing graphic organisers in the case of the metacognitive (with planning) group or mere provision of pre-task planning opportunities in the case of the non-metacognitive (with planning) group would help the two groups to acquire the required knowledge about planning and move toward the mastery of higher level content with more confidence and understanding. This hypothesis was grounded in Skehan and Foster's (2001) assertion that pre-task planning reduces the cognitive demands which are placed on learners during the process of text generation. It appears, however, that the participants under the no planning condition, although instructed to immediately start composition without planning, were engaged in an online planning during task execution. Thus, the impact of the pre-task planning in the metacognitive condition, although hypothesised to make the writing task cognitively less demanding, based on Skehan and Foster's (2001) Limited Attentional Capacity Model, did not exceed that of the metacognitive (without planning) condition. The metacognitive (without planning) group appeared to have obtained even greater automaticity in simultaneously managing their conscientious planning and the formulation process in the delayed post-test, which resulted in their superior performance in comparison with Group -M-P and the control group. The next planning and no-planning comparison between Group -M+P and Group -M-P in the immediate and delayed post-tests did not uncover any significant results either. This lack of highly skilled performance of planning to no planning groups shows that the results did not reflect any global effect of pre-task planning on the overall quality of written texts.

A tentative explanation may be that within-task planning can be as important as pre-task planning. In this study, the 10-minute planning time was associated with reduced composing time (35 minutes). Accordingly, the planning groups had more time to plan but were pressured to complete the essay in the time allotted. On the other hand, the no-planning groups were also pressured in the sense that they had 45 minutes to formulate, execute and monitor the task rapidly and had limited opportunity to simultaneously attend to content and form. Furthermore, writing is a recursive process in contrast to the more linear speaking process. One cannot claim with certainty that the writers in the planning groups would closely follow their planning notes. As the text develops, it is likely that they would react differently to their pre-planned brainstorming notes and engage in additional within-task planning. This way, the effects of pre-task planning may have been obscured.

These findings are in contrast to the results of a number of L1 (e.g. Kellogg, 1988, 1990) and L2 writing studies (e.g. Mahnam & Nejadansari, 2012; Shi, 1998) which found significant effects of pre-task planning on writers' overall text quality. In their study, De La Paz and Graham (1997) also indicated that providing explicit instruction in planning prior to composing resulted in students producing more complete and qualitatively improved essays. Research on planning strategy instruction has demonstrated that planning is a powerful approach which helps students to improve the quality of their writing (e.g. Graham & Perin, 2006). Graham and Harris (2007) assumed that the goal of explicit instruction of planning strategies is to promote students' effective, independent, flexible, and motivated use of those strategies. This mastery of the planning processes, they asserted, can later result in students' valuing and continuously using them. They argued that some strategies, like brainstorming, have wider applicability in the sense that they can be applied in a wide variety of writing tasks. These impacts were not, however, observed when a comparison between planning and no planning conditions was made in the present study.

The difference between previous investigations and the present study might be explained in terms of the discourse requirements of argumentative essays. In a large number of previous studies on the impact of planning on L2 writing, participants were asked to write narrative texts, whereas in the present research project, participants were required to produce argumentative texts. Unlike argumentative genre, which has been shown to

impose more attentional demands on writers, written narratives have been demonstrated to be less difficult because of learners' more exposure and experience with this genre (M. D. Johnson et al., 2012; Kellogg, 1996). Given the results of the current study, it is likely that the impact of pre-task planning was moderated by the cognitive demands of the argumentative writing process.

5.2.2. Research Question 2

Effects of metacognitive strategy instruction and pre-task planning on the accuracy of L2 learners' written production

With respect to the second research question, as illustrated in Figure 5.2, findings indicate that the five groups displayed highly similar values in terms of the accuracy of clauses in pre-test, immediate post-test and delayed post-test. When within-group differences across time were measured for the production of texts which contained fewer error-free clauses, only Group +M+P showed a significant progress from the pre-test to the delayed post-test.

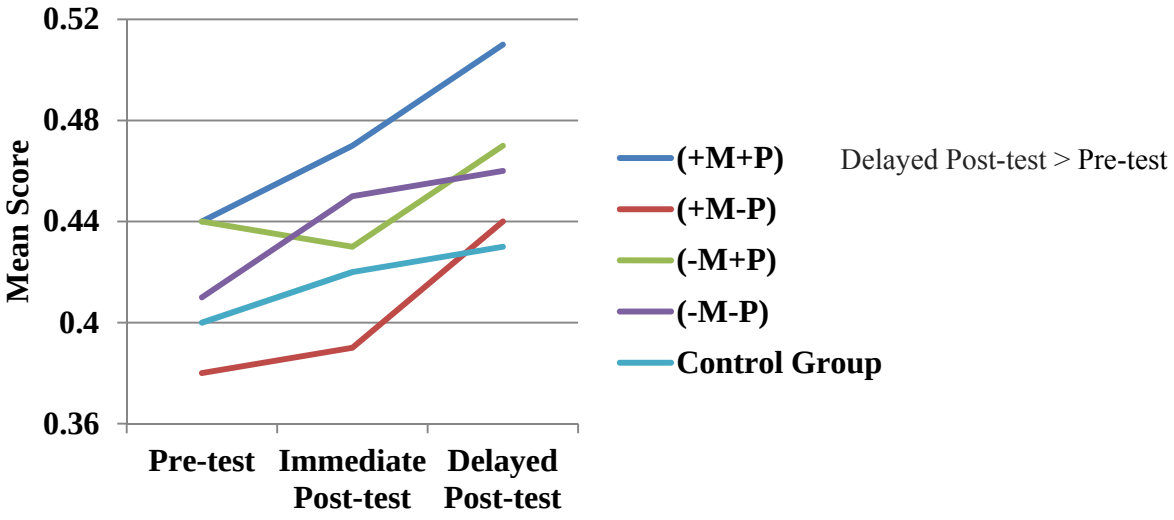


Figure 5.2 Accuracy (Error-free clauses)

Note. M: Metacognitive instruction, P: Planning

As shown in Figure 5.3, there were not only non-significant differences among groups in terms of the accuracy of verb forms in the pre-test, and the two post-tests, but also there occurred no significant progress over time in any of the groups.

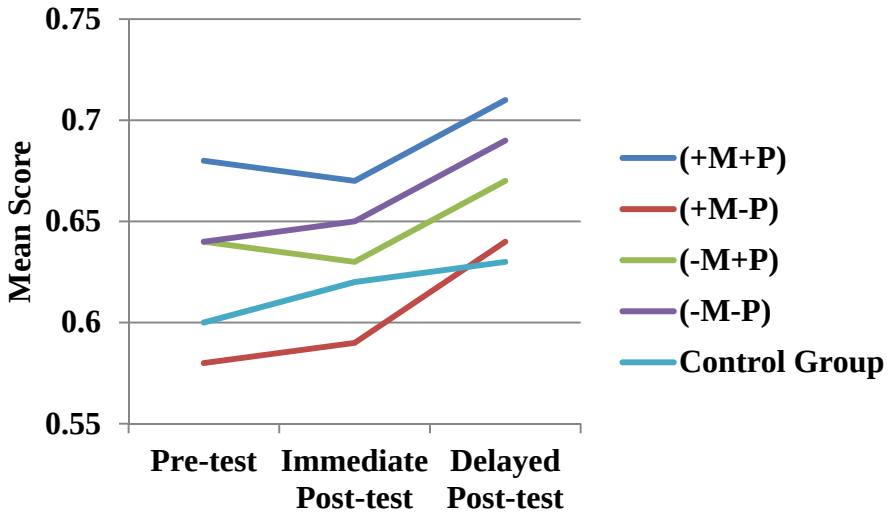


Figure 5.3 Accuracy (Error-free verbs)

Note. M: Metacognitive instruction, P: Planning

Thus, the answer to the second research question is negative based on the findings that there were non-significant differences between the metacognitive and non-metacognitive conditions and between the planning and no planning conditions as regards the two measures of accuracy. Only when combined with pre-task planning did the metacognitive instruction proved to bring about significant result in terms of error-free clauses from the pre-test to the delayed post-test.

- ***Effects of metacognitive instruction on accuracy***

Schmidt (1990) makes the assumption that learners have limited attentional capacities available to them, thus they choose to devote attention to one area at the expense of other areas, particularly when choosing to attend to form and/or meaning. The same limited attentional capacity argument applies to the present study in the sense that the writing tasks might have placed so many demands on participants in all groups that their attention was pushed away from form to meaning. Given that argumentative writing requires more complex grammatical structures, it is likely that the participants could infrequently attend to their linguistic repertoire and, as a result, the accuracy of the verb forms and clauses in their output was not enhanced in a marked way. In a similar vein, as Manchón and Roca de Larios (2007, p. 555) claim, the complex nature of writing and the L2 writers' limited attentional capacity requires them to handle various sets of competing ideational, linguistic, and rhetorical demands of composition. Overall, previous literature consistently confirms

that due to the limitations in their working memory capacity, L2 learners have difficulty simultaneously attending to form and meaning (DeKeyser, Salaberry, Robinson, & Harrington, 2002; Robinson, 2003; Skehan, 1998; VanPatten, 1996, 2002).

Another relevant factor in discussing these findings is that training in no way intended to assist the participants in their development of accuracy, which is an aspect of general proficiency, and an outcome of long-term exposure and use rather than metacognitive strategy instruction. The non-significant results obtained for the accuracy measures relate more clearly to the nature of instruction which, as Foster and Skehan (1999) have posited, challenges learners' attention and affects the priorities in their performance. The participants in the two metacognitive groups were explicitly and systematically instructed on strategies for planning, drafting, and revising their essays. It was assumed that once they mastered these strategies, they would learn to use them more flexibly and independently. They were not, however, directed to use certain linguistic features. Yet, during a writing task performed within time constraints, one would expect the learners to be reliant primarily on their implicit knowledge to complete a task (Ellis, 2005). Increased accuracy might be expected only when the participants had time to monitor their writing, check for grammatical and/or lexical errors, and make use of explicit knowledge. Given the limited time-frame for writing in the current study, this was unlikely.

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

In order to determine the true effects of pre-task planning versus no planning in both metacognitively and non-metacognitively instructed conditions, the amount of time spent on task was manipulated. The planning groups in both conditions were given 10 minutes as the planning time and 35 minutes as the transcription time in the formulation process. The no-planning groups, however, were instructed not to plan, but to write immediately and continuously for 45 minutes. Thus, although given no longer time in total, the planning groups were required to spend 10 minutes for planning before embarking on the writing task itself. Such planning was guided for Group +M+P, in the sense that the participants were instructed how to prepare and what to plan for the upcoming task. All participants were therefore given a time limit for performing the task, which in turn pressurised them almost equally to make a limited use of within-task planning. It was hypothesised that the extent to which there was an opportunity for strategic planning, which I have called pre-task planning in the present study, would free the working

memory resources from the multiple demands of the writing process and would positively influence the performance. Nonetheless, the non-significant results obtained in comparing the planning groups with the no planning groups may be that they were not advised to focus on specific linguistic forms that they would need when performing the task. They had to rely largely on their own linguistic and non-linguistic resources in order to complete the written task.

At the same time, it was hypothesised that if planning was contemplated as a pre-task activity, it would be produced as a framework in the form of written outlined plans and would function as a set of guidelines. This inclusion of planning would accordingly assist participants to set goals and to try out means to achieve those goals. Thus, it was expected that planning would provide additional attentional resources for participants to attend to form during the task performance. This, however, resulted in better performance for only Group +M+P when the error-free clauses were examined from the pre-test to the delayed post-test. This finding shows that when these participants repeated writing several tasks of the same type, they started not to focus exclusively on meaning, but they may have learnt to attend to form as well. This was also supported by Graham and Harris (2007) who emphasised the predictability ingredient of planning in students' writing routine in the classroom. In order to increase the likelihood that students would plan their writing, Graham and Harris asserted that it would behoove teachers to create a supportive writing environment where students can value and make effective use of planning. In accordance, they rightly argued that pre-task planning can function as an external memory where ideas are stored and are "readily available for inspection, reflection, and reconceptualisation" (p. 120). Scrutinising skills such as these can be particularly advantageous, as the written plan can minimise the need of online planning by freeing the resources which are required in other attention-demanding processes, such as transforming ideas into well-written sentences.

With respect to the provision of pre-task planning time and its effect on accuracy over time, the results, however, show that the impact was not true for the non-metacognitive (with planning) group. Although they were allowed to pre-write before the actual composition, they had not received explicit instruction in using each of the pre-task planning techniques. Even though the time on task was held constant for all the participants, rather than simply giving 10 minutes in advance of writing task for the

planning, participants in Group +M+P were provided with the explicit instruction on how to make a plan in the first session. The significance of providing sufficient guidance in planning has been acknowledged to be positive in writing contexts (e.g. Ellis, 2003; Ojima, 2006). In addition to providing guidance in how to plan, Ojima (2006) indicated that learners needed practice in familiarising themselves with planning strategies. For Group +M+P, the explicit instruction in the analysis of metacognitive strategies was combined with scaffolded practice in the use of those strategies. This suggests that only when effective pre-task planning activities are coupled with the use of metacognitive strategies do they have the potential to improve accuracy of writing over time.

Previous studies of pre-task planning have yielded mixed results where accuracy was concerned. With respect to accuracy in L2 writing, the finding of the current study is at odds with three studies conducted in Iranian EFL contexts (Haghverdi et al., 2013; Khomeijani Farahani & Meraji, 2011; Rahimpour & Nariman-Jahan, 2011). In a study investigating the effects of strategic, within-task, and no planning on the accuracy of both male and female learners' narrative writing, Haghverdi et al. (2013) found that strategic planning yielded the most accurate performance. They asserted that planning could ease the cognitive processing load during task completion and enabled the learners to attend to form as well as meaning. Similarly, Rahimpour and Nariman-Jahan (2011) revealed that pre-task planning resulted in the production of more accurate narrative performance for high proficiency learners. Their study, however, did not show any significant improvement in accuracy for low proficiency learners. The explanation they provided was that the former group used the planning time to focus on the form, whereas the latter group when provided with time to plan focused on the content of their linguistic output.

However, the findings in the present study are in line with the results of other studies conducted both in Iranian EFL contexts (e.g. Piri, Barati, & Ketabi (2012); Rahimpour & Nariman-Jahan (2011) (for low proficiency learners); Rahimpour & Safarie (2011) and other L2 contexts (e.g. Ellis & Yuan, 2004; Ojima, 2006; Ortega, 1999; Wendel, 1997). These studies have demonstrated that pre-task planning has no significant effect on accuracy. As a representative example of these findings, Ellis and Yuan (2004) reported that pre-task planning did not improve accuracy, but on-line planning resulted in increased accuracy. Likewise, Ojima (2006) used a case study approach to examine the effect of concept mapping, which is a form of pre-task planning. Similar to Ellis and Yuan (2004),

Ojima (2006), in his process analysis of the learner's planning, showed that pre-task planning activities are directly connected to learners' written production at conceptual and textual levels. Additionally, the results of the text analysis showed that pre-task planning, particularly when accompanied with the provision of guidance in how to plan, improved lexical complexity, although not accuracy.

In conclusion, compared with the previous research studies, the findings of the present study showed that neither pre-task planning nor metacognitive instruction in isolation played significant roles in reducing the cognitive load the task placed on participants and in helping the participants to attend to explicit syntactic knowledge. Nevertheless, the current study found an increase in the accuracy of clauses when measuring the progress of the metacognitive (with planning) condition from the pre-test to the delayed post-test. This suggests that only when pre-task planning activities are coupled with metacognitive instruction do they return significant results over time.

5.2.3. Research Question 3

Effects of metacognitive strategy instruction and pre-task planning on the complexity of L2 learners' written production

Regarding the measure of complexity, the study findings can be discussed in terms of syntactic complexity, syntactic variety, and lexical variety. According to Figure 5.4, no significant differences were found among groups in terms of the syntactic complexity of production in either pre-test, or in the two post-tests. The control group appeared to perform differently from other groups in the delayed post-test. They appeared to produce more errors but because the increments were small, the difference was inconsequential.

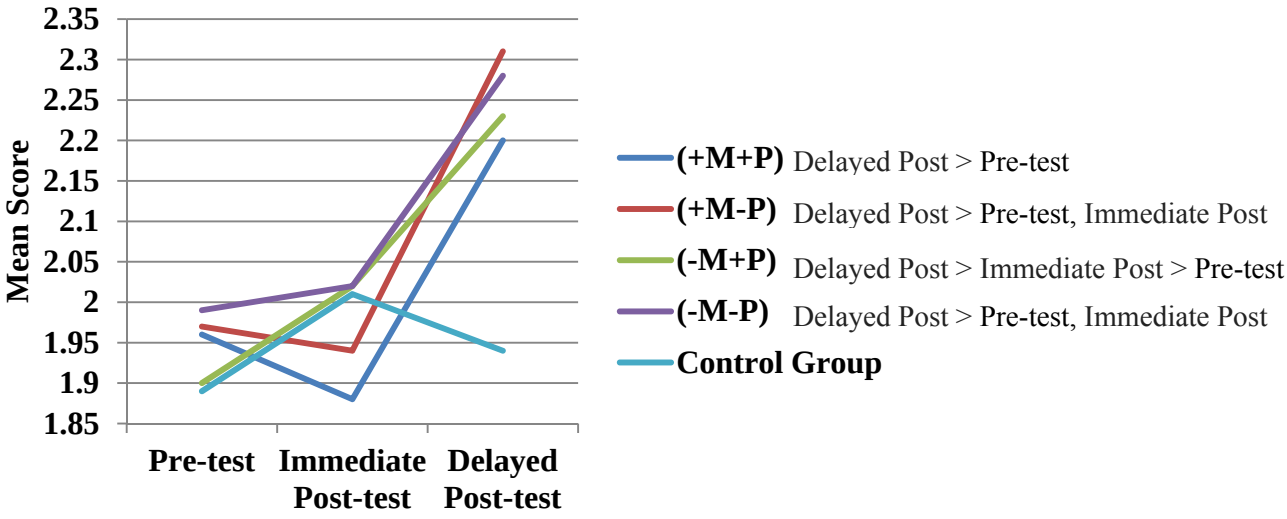


Figure 5.4 Complexity (Syntactic complexity)

Note. M: Metacognitive instruction, P: Planning

The same results were obtained for the comparisons between groups with regard to syntactic variety (see Figure 5.5). No differences between groups were detected in the three task performances. Thus the answer to this research question with regard to the first two measures of complexity when compared between groups was negative.

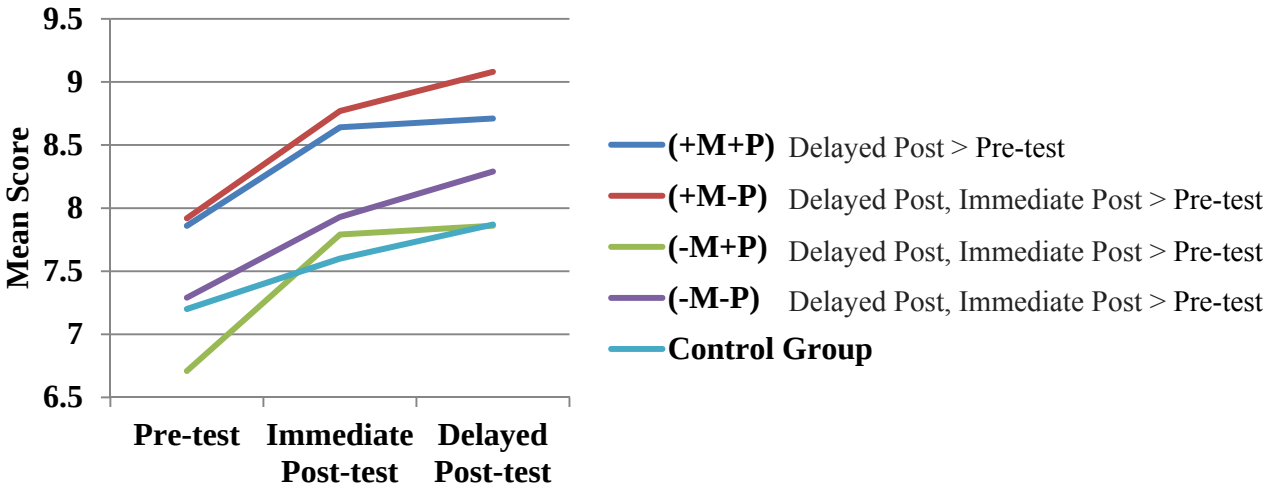


Figure 5.5 Complexity (Syntactic variety)

Note. M: Metacognitive instruction, P: Planning

Despite the lack of significant differences between groups, when within-group comparisons over time were taken into account for these two measures of complexity, interesting results were obtained. In terms of syntactic complexity (see Figure 5.4), Group +M+P progressed producing more syntactically complex texts only in the delayed post-test compared to the pre-test. All other treatment groups progressed in the delayed post-test compared to both pre-test and immediate post-test. With regard to syntactic variety (see Figure 5.5), Group +M+P progressed in all three comparisons over time. Meanwhile, all other treatment groups progressed in both immediate and delayed post-tests compared to the pre-test. Given that all tasks during the treatment, and pre- and post-tests were similar in type, there was a possibility of a practice effect, and thus the changes were not necessarily due to the provision of planning time or metacognitive instruction. Therefore, the answer to this research question with regard to these two measures of complexity when compared within groups is not affirmative either.

The answer to the third research question with regard to lexical variety can be partially confirmed. In the immediate post-test, Group +M+P produced more lexically varied writing than Group -M-P and the control group. In this time phase, Group +M-P also outperformed the control group. This superior performance was not, however, completely observed along the delayed post-test in the sense that the beneficial effects of metacognitive instruction on lexical variety which were observed in the immediate post-test were not maintained in the delayed post-test. Only Group +M+P showed a tendency to outperform Group -M-P.

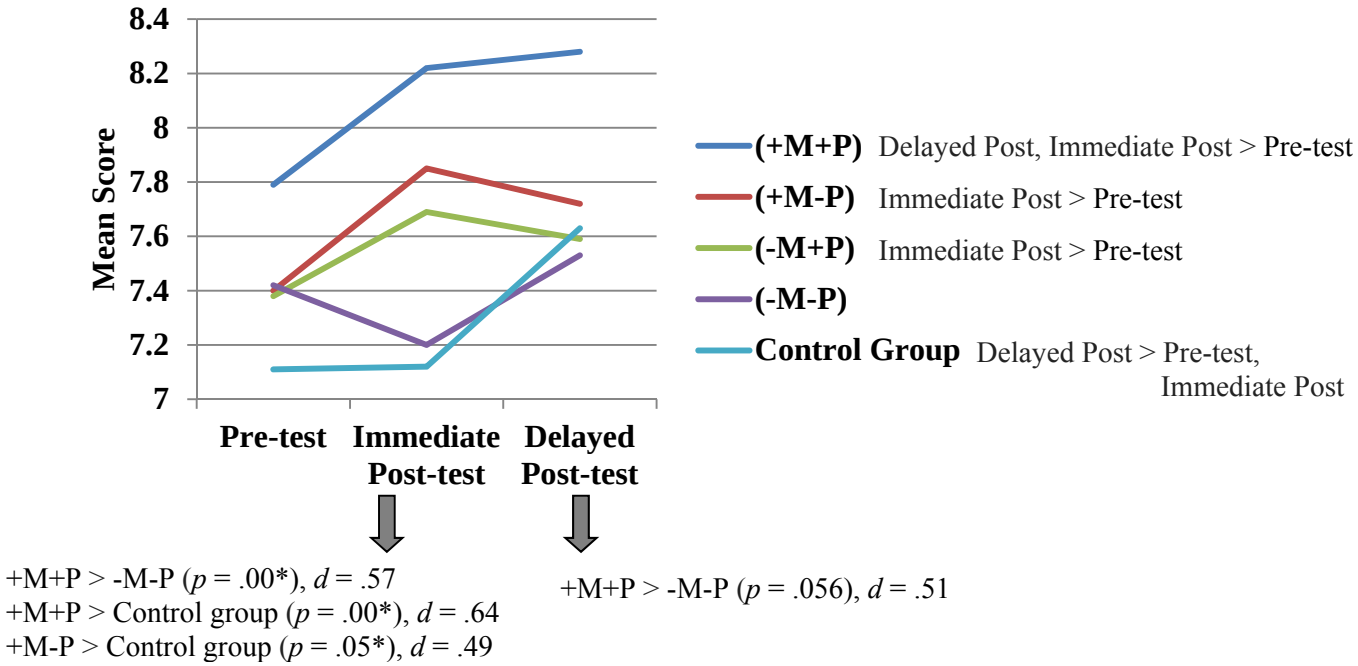


Figure 5.6 Complexity (Lexical variety)

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

Both metacognitive instruction and pre-task planning appeared to have helped when the progress over time was taken into consideration. Not only did the two metacognitively trained groups progress over time, but also Group -M+P seemed to have been advantaged by the planning in producing texts which were more lexically varied in the immediate post-test compared to the pre-test. Even the control group progressed in terms of producing written texts with more lexically varied sentences. The only group which did not show any significant change over time was -M-P.

Before separately examining the effects of metacognitive instruction and planning on complexity, the reasons behind the progress of the control group over time can be presented. One possible explanation is that, given the nature of argumentative writing, the control group found the opportunity to reflect on propositional content, despite the fact that they were pressured to write quickly. Thus, even the control group was pushed to find sufficient time for lexical searching while writing the task, and prioritised this aspect of verbal processing. This may also be due to the vocabulary that they picked up in the course of their attendance in other classes. This is to say that the progress observed here

was not necessarily a pedagogical by-product of writing courses. It can be argued that lexical variety can be improved as a by-product of the involvement of the language learner in analysing modes of producing a text other than the act of writing itself. It is also safe to argue that linguistic input per se, irrespective of the type of course, factors in developing lexical complexity in the individual learners' writing performance. The raw fact remains that any language learner is basically a language user in the first place. Knowledge and information are dynamically acquired and frequently updated in classrooms and other structured training environments (Everson & Tobias, 2001). This understanding of the language learner specifically applies to the English language learners in the current study who aimed at sitting IELTS exams. They were mostly experienced adults with an adequate educational background. In this light, these language learners could consciously and actively draw upon and import the lexical knowledge that they acquired from other means of exposure to English language. Needless to say the possibility of gaining lexical knowledge goes beyond the physical boundaries of language classrooms. Popular means of entertainment and communication such as satellite, internet, books and even music can be deemed as textual entities to impart lexical knowledge to their users. According to Wikipedia, "Communications in Iran" (2009) despite restrictions and censorships in Iran, there are over 45 million phone users, over 18 million private internet users and over 23 million public/internet cafe users (out of the 75 million population), making it primary in the Middle East, particularly in terms of the number of internet daily users.

- ***Effects of metacognitive instruction on complexity***

The two measures of syntactic complexity and syntactic variety were not shown to be significantly different between groups, but the measure of lexical variety was. Both immediate and delayed post-test results for the metacognitive (with planning) group compared to the non-metacognitive (without planning) group can be attributed to the nature of instruction. There are four possible explanations as to why metacognitive training, particularly when strengthened by the pre-task planning opportunity prior to composing activities, allowed participants to produce greater lexical variety. The first reason may be that the metacognitive treatment involved explicitly teaching participants the strategies for planning, monitoring and evaluating. The within-class activities included procedures on how to plan in advance that included brainstorming ideas and content, and organising ideas through the use of semantic webs. These mind maps might have helped them to quickly get down a lot of related words and ideas on a single piece of paper. This

measure of complexity (using Guiraud index of lexical richness) improved incrementally and, as time passed, it was the metacognitive group which reaped the benefits.

Secondly, among the components of metacognition are the abilities to judge the cognitive demands of a particular task, and the assessment of one's progress both during and after the performance (Flavell, 1979). Likewise, the metacognitive groups who had apparently learned to acquire this awareness and were trained to be in control of their metacognitive behaviours were at a distinct advantage. Both Group +M+P who had the planning time and Group +M-P who were instructed to start writing immediately and without pre-task planning were trained to metacognitively approach an argumentative writing task. This encouragement might have prompted an automatic lexical retrieval, which in turn triggered them to use more complex and sophisticated lexical items.

Thirdly, genre knowledge may explain why participants in the metacognitive (with planning) group appeared to be able to retrieve a wider variety of words during the composing process in comparison to the participants in the non-metacognitive (without planning) group. The curricula of writing courses for the latter group focused on a familiarity with and an analysis of four typical discoursal modes in essay writing (i.e. narrative, descriptive, expository, and argumentative), but practiced the composition of argumentative essay in each of the four writing practices during the treatment. Only in the last session of the treatment, both the instruction and the practice focused on argumentative essays. Thus, participants were attempting to use their knowledge of a familiar genre in the first three sessions to complete a writing task which required them to compose in an unfamiliar genre, i.e. argumentative essay. However, there is a possibility that the complexity of the language of argumentation was constraining the development of their chain of ideas. This is also one of the possible explanations which M. D. Johnson et al. (2012) provide for finding no impact of pre-task planning on the grammatical and lexical complexity of L2 writers' texts. They suggest that many of their participants composed comparison-contrast essays in response to the assigned argumentative prompts, mainly because the students' writing courses had focused on comparison-contrast essays.

A fourth possible explanation is in terms of Robinson's (2001a, 2001b, 2003) Cognition Hypothesis which makes clear predictions about the impact of task complexity. He posits that higher cognitive task demands can lead to greater complexity. These predictions were

only confirmed for lexical complexity, but not for the two measures of complexity. Similarly, a number of previous studies have shown that the effect of task complexity is more evident on lexical diversity (e.g. N. H. De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Robinson, 2001b). Thus, in line with these studies, it was clear why in the argumentative essays, which are considered complex tasks, participants in the metacognitive groups used more diverse language.

It can be argued that the progress of all groups over time in the three measures of complexity is due to the fact that all participants were required to write in the same genre of argumentative writing. As Kormos (2011) has argued, genre affects the lexical and syntactic range of expressions that learners use. Argumentative genre demands elaborated content and advanced vocabulary and grammatical structures. Therefore attending to the lexical properties of their message was a necessity for participants in all conditions. Accordingly, it is possible that the requirements of essay writing might have entailed participants to gradually learn to pay more attention to lexical and syntactic encoding in the composing process. Even at times, heeding linguistic properties of the text has been reported to be the cause of learners experiencing difficulties in devoting sufficient resources to the discourse structure and global organisation of the text (McCutchen, 1996). These macro-level features of texts including idea development, sufficiency, and organisation of information were, however, not investigated as potential signs of enhanced quality in this study.

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

It is important to stress that of the three main systems in Kellogg's (1996) model of working memory in writing (i.e. formulation, execution, and monitoring), formulation is theorised to place the greatest demands on working memory capacity. According to Kellogg, pre-task planning can reduce such demands and result in an increased use of grammatically complex forms. In the present study, the participants in the no planning groups were encouraged to write continuously without pre-task preparation. They were not involved in deliberate planning during the formulation process. Therefore, it was hypothesised that this would be a detriment to the complexity of their performance. Yet, the results showed that pre-task planning did not help the planning groups to outperform the no planning groups in their production of texts with enhanced syntactic complexity and syntactic variety. The reason for this lack of significant difference might be related to the

fact that, as Skehan and Foster (2001) have also hypothesised, when participants were faced with the cognitively demanding tasks of academic writing, they allocated their attentional capacity to conveying meaning first, and the linguistic complexity and accuracy next.

Previous studies have shown mixed results in the case of grammatical complexity. The results in the present study with regard to syntactic complexity are contrary to the findings of Ellis and Yuan's (2004, 2005) studies of the impact of planning on L2 written narrative. They suggested that the specified period of time prior to composing significantly impacted syntactic complexity. Yet, the findings confirm the results of prior recent research conducted on the role of pre-task planning on the complexity of L2 writing in Iranian EFL context (e.g. Piri et al., 2012; Rahimpour & Safarie, 2011). Piri et al. (2012) pointed out two reasons why pre-task planning was not of much value in helping the Iranian EFL learners to produce a complex narrative text. First, they believed the narrative task in their study was not challenging enough, and thus the pre-task planning did not help them to enhance writing in terms of complexity. Secondly, they assumed that the learners had no familiarity with planning strategies and were not aware of the effective self-regulatory strategies like planning, monitoring, evaluating and revising which are the strategies applied by skilled writers. These explanations do not, however, apply to the findings in the present study. The argumentative writing tasks in this study were challenging enough and even the metacognitively trained groups, who received instruction on how to plan prior to writing, did not produce writing that was more syntactically varied or syntactically complex than that of the non-metacognitive groups.

Although no consequential results as regards the lexical complexity were obtained when planning and no planning groups were compared in three time phases, interesting insights can be gained from a comparison of their performance over time. Participants in both metacognitive groups were provided with some extended instructional support in carrying out pre-task planning. Group +M+P, in particular, progressed from the pre-test to both immediate and delayed post-test. My resulting observation is that this planning opportunity helped reduce the cognitive demands placed on the participants during the task of writing in a way which enabled them to engage in the retrieval of a wider variety of words. Group +M-P, even though instructed how to plan, was deprived of the opportunity to allocate some time to planning. They progressed only from the pre-test to the immediate

post-test. According to Ellis (2012), detailed planning, which involves the provision of guidance to learners on how to use the planning time to consider the grammar, vocabulary and organisation of their outcome can lead to more complex and at times more accurate language use. This is because when a task is divided into subtasks including planning, it is simplified by means of setting goals and planning the content and intention of the message (Manchón & Roca de Larios, 2007). It is possible that if Group +M-P could also draft a predetermined framework, they would have progressed like Group +M+P.

In addition to the two metacognitively trained groups, it is interesting that Group -M+P and Group -M-P also differed in their progress: Group -M-P showed no progress. It is likely that the reason why the non-metacognitive (with planning) group progressed in producing more lexically varied language over time was that they employed the pre-task planning to compensate the cognitive load required of the task and access their lexical knowledge. According to Ellis (2005), planning helps learners to overcome their limited working memory capacity. This result can also be plainly explained in terms of Skehan and Foster's (2001) Limited Attentional Capacity Model which postulates limited attention and memory resources for learners. When dealing with more cognitively demanding tasks in particular, they argued that learners are forced to concentrate on the content of the message to the disadvantage of accurate linguistic forms. As a result, they assume a trade-off between complexity and accuracy, and this was also observed in the present study. This trade-off explains why richer and more elaborate lexicalisation was achieved to the detriment of accuracy. Possibly, participants were engaged in planning for content rather than language for which they required to identify lexical structures rather than syntactic features. This suggests that unlike the results of Ong and Zhang's (2010) study, where they claim that pre-task planning impedes the lexical complexity of L2 writers' texts, planning can in fact facilitate enhanced lexical variation in participants' writing over time.

To explain the contrast between the findings of this study and the findings of previous studies, two explanations are suggested. The first explanation suggests a fundamental difference between speaking and writing. The results obtained in studies of oral language production have shown that pre-task planning helps to reduce the cognitive demands placed on learners' working memory resources during task execution and results in greater complexity in their language production (e.g. Crookes, 1989; Foster & Skehan, 1996;

Mehnert, 1998; Ortega, 1999; Skehan & Foster, 1999; Wigglesworth, 1997; Yuan & Ellis, 2003). In all these studies, planning is partitioned from the process of oral execution process. Ong and Zhang (2010), however, speculate that in written language production, learners in the planning condition still continue to plan during the task execution and thus the quantity of planning (pre-task plus online) exceeds that of learners who are encouraged to write continuously without planning. It is likely that the same explanation applies here as well. Due to the recursive nature of writing, further planning may occur during online planning and monitoring, and thus the impact of pre-task planning is likely to be obscured.

The second explanation is that the measure of lexical variety used in the present study was calculated via Guiraud index of lexical richness, which is more sensitive to text length. In many planning studies in L2 writing, however, the mean segmental type-token ratio (MSTTR) was calculated by dividing the total number of different words by the total number of words in the segment. Yet, a number of these studies like Ellis and Yuan (2004) acknowledge the limitation of MSTTR measure which is mainly its insensitivity to "repetition of words beyond the boundary of their own segment" (Malvern & Richards 2002, p. 88). Some of these studies reported to lack the software needed to compute other more valid measures of lexical complexity, including Guiraud index and/or D-values.

In sum, the major claim of Skehan and Foster's (2001) Limited Attentional Capacity Model which hypothesises a trade-off effect between linguistic form and meaning and between accuracy and complexity is supported in this study. However, this model's premise which postulates an increase in cognitive task complexity will hinder the development of accuracy and complexity is not valid in terms of the lexical complexity. On the contrary, it was found that when the need for attentional resources increased during the complex task execution, these resources were mainly focused on the control of lexical forms in the metacognitive groups. This is in line with Robinson's (2001a, 2001b, 2003) Cognition Hypothesis which hypothesises that more complex tasks trigger the use of more complex syntactic structures and lexical forms, although no effects on syntactic complexity were observed.

Summary

The empirical evidence presented in this chapter returned interesting results. First, given the paucity of task-based research on written language production, the current study

explored the effects of manipulating both the provision of pre-task planning and the instruction of metacognitive strategies. In accomplishing this aim, quantitative measures of writing proficiency, complexity, and accuracy were statistically compared across four experimental groups and a control group. Due to the combination of pre-task planning and metacognitive instruction employed in this study, it was at times difficult to make a comparison between the current study and previous investigations that have been conducted on written task production. Overall, the findings showed that teaching metacognitive strategies can result in marked improvement in the lexical variety of written performance on the one hand and the general performance of argumentative essay development on the other. The results therefore support the research encouraging the significance of metacognitive training for writing (e.g. Hacker et al., 2009; Harris et al., 2010). Furthermore, what is evident from the results is that these pre-task planning activities of goal setting and brainstorming per se may not have any measurable impact on features of learners' texts. Rather, when coupled with the knowledge and mastery of other metacognitive strategies, they provide firm evidence in favour of metacognitive strategy instruction.

Secondly, the findings of this study are in contrast to other studies which have been conducted on planning. On the one hand, some studies such as Ong and Zhang (2010) have suggested that pre-task planning impedes the lexical complexity of L2 writers' texts. On the other hand, other studies such as Ellis and Yuan (2004) found large positive effects of planning on the syntactic and lexical variety of L2 written narratives. However, the results of this study support those of Johnson et al.'s (2012) who found no impact of pre-task planning on the grammatical complexity of L2 writers' texts.

Thirdly, it is worth restating that this study provides a plausible explanation for information processing theories which claim that human beings possess a limited processing capacity and are therefore not able to attend fully to all aspects of a task (Schmidt, 2001; Skehan, 1998). In addition, the results lend support to the predictions of Skehan and Foster's (2001) Limited Attentional Capacity with respect to learners prioritising the processing of meaning over processing of language form. This trade-off, which was observed between accuracy and lexical complexity in this study, suggested that accuracy and syntactic complexity was compromised as a result of improved lexical complexity.

CHAPTER 6. RESULTS AND DISCUSSION: RESEARCH QUESTION 4

Utilising mixed methods in this research was considered an appropriate methodology for two reasons. First of all, since the primary aim of the study was to explore the impact of metacognitive instruction and pre-task planning on L2 learners' argumentative writing ability, the study needed to analyse participants' writing tasks in the pre-test and post-tests quantitatively. Accordingly, such data were examined holistically (measuring general writing proficiency) and analytically (measuring accuracy and complexity). Secondly, with the same aim in mind, it needed to explore participants' perceptions of their application of strategies more specifically to gauge their perceptions towards the absence/presence as well as the frequency of occurrence of each strategy.

This chapter aims to address the fourth research question:

> 4. What metacognitive strategies do L2 learners perceive in their actual writing performance?

Data were collected through think-aloud protocols, and Metacognitive Strategy Questionnaires (MSQ). Thus, in order to add more richness and detail to the discussion of findings when answering the questions of 'why' and 'how' the purely statistical results appeared as they did, these two additional types of data were collected and analysed to function as more meaningful adjuncts to the quantitative side of the research. Due to the complementary nature of these analyses, as well as the amount of quantitative written data collected to answer the first three research questions, the results and discussion of these two instruments are presented together in this chapter.

6.1. Think-aloud protocols

According to A. L. Brown, Bransford, Ferrara, and Campione (1983), metacognitive knowledge is statable. Ever since Hayes and Flower (1980) introduced the think-aloud methodology to writing, it has been extensively used to access writers' cognitions (e.g. Bereiter & Scardamalia, 1987; Kellogg, 1987). Hacker et al. (2009) conceptualise writing

as a metacognitive process in which an act of meaning production occurs under the direction of metacognitive monitoring and control processes. They assert that investigating this moment-by-moment production via think-aloud protocols enables researchers to gain an understanding of both thinking and thinking about thinking (p. 170). As qualitative research is concerned with "studying things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them" (Denzin & Lincoln, 1994, p. 2), the data obtained from the think-alouds in the present study were analysed qualitatively. This aimed to provide insights into actual writing performance by exploring the more subtle aspects of writing behaviour. In Study 2, some volunteer participants were asked to reflect on what they knew and did while writing in order to help explore their metacognitive awareness. The results reported in this section are based on six volunteer participants from whom think-aloud protocols were collected over four writing practices. They were one male and one female participant in Group +M+P, one female in Group +M-P, one male in Group -M+P and two males in Group -M-P. The interpretations of such analysis, however, are not intended to be generalisable beyond the sampling of participants on whose thoughts detailed analyses were conducted.

6.1.1. Protocol coding

In Study 2, twenty four protocols collected from six participants during four instruction sessions were analysed to show their use of strategies during the writing practices. Ericsson and Simon's (1993) model suggests that there are no performance differences between subjects who think aloud during task completion and those who complete the same tasks silently. Yet, the think-aloud participants' written texts were not included in the quantitative textual analysis and were investigated with the sole purpose of gaining insights into their cognitive processes. This was done to eliminate any likelihood of reactivity, or the fact that the act of thinking aloud may potentially trigger changes in learners' cognitive processes while performing the task (Stratman & Hamp-Lyons, 1994, p. 90). The correspondence between thinking aloud and its impact on learning performances is contested, since for instance, Bannert and Mengelkamp (2008) claim that learning performances are not influenced by thinking aloud, and Fox, Ericsson, and Best (2011) maintain that although all verbal reporting procedures increase the time required to complete a task, they do not alter task performance. They therefore conclude that thinking aloud is a nonreactive measuring method. However, to eliminate the possibility of

reactivity, the think-aloud participants' written data were excluded from the analysis conducted to answer the first three research questions.

A series of codes were utilised to identify potential themes and trends and to categorise the data collected. The coding scheme for this study was adapted from Flavell, Miller and Miller's (2002) division of metacognition into metacognitive knowledge and metacognitive monitoring and self-regulation. Metacognitive knowledge refers to an individual's declarative knowledge of learning strategies, person, and task characteristics. Metacognitive regulatory skills refer to control, monitoring, and self-regulation activities that take place during learning and problems solving tasks. In this research, the main focus of instruction was the second component of metacognition, i.e. teaching metacognitive regulatory skills to participants. One of the differences between the categories used in studies such as Schellings (2011) and Schellings et al. (2012) and the categories used in the current study in both designing the lesson plans and devising the instruments to measure the metacognitive strategies is that the categories in the current study are theory-driven, and derived from Flavell's (1979) original three-fold categorisation of metacognitive regulation activities: planning, monitoring, and evaluating. In the two studies mentioned above, however, the categories are data-driven, meaning that an empirically driven list of metacognitive categories is presented based on an overview of other measures of metacognitive activities. These are more specifically think-aloud data, or may be obtained following interview-based research.

Table 6.1 illustrates the categories used for coding participants' focus during the think-aloud, along with brief definitions and some sample fragments from the data.

Table 6.1 Metacognitive strategy categories in think-aloud protocols

Focus		Definitions	Samples
Metacognitive Regulatory Skills	1. Planning	Statements or questions about what will or should happen next.	"Now I need to think of the body paragraph."
	2. Considering the audience	Statements or questions about sensitivity to the audience.	"I hope the reader will not be lost."
	3. Monitoring	Statements or questions noting progress, or lack thereof, on the writing task.	"I'm going too slowly." "I think I have been using the word 'support' too often here." "Am I going the right way in how I am discussing my view?" "This is my point; I need to stick to my point."
	4. Evaluating	Statements or questions concerning conclusions on ending the writing task – regarding the product, ability, or the experience of doing the task.	"This is my best one so far!"; "Ambiguous.. I need stronger evidence to present my idea."; "It is a good start, very relevant introduction."
Metacognitive Knowledge	1. Person	Statements about self and what kinds of tasks one is good at.	"I usually write less than the word limit."
	2. Task	Statements about the tasks one undertakes.	"This is an argumentative writing."
	3. Strategy	Statements about strategies and procedures available for achieving the goal.	"I need at least four paragraphs; 1 introduction, 2 body and 1 conclusion."

As shown above, the statements coded as metacognitive regulatory skills reflect what Flavell et al. (2002) term metacognitive monitoring and self-regulation, and others (e.g. Jacobs & Paris, 1987; Kluwe, 1987) describe as consisting of planning (i.e. the selection of appropriate strategies and the allocation of resources), monitoring (i.e. one's on-line awareness of comprehension and task performance) and evaluation (i.e. appraising the products and regulatory processes of one's learning). The second component of

metacognition is metacognitive knowledge. This provides information about writers' mental models of the writing process in the three areas of metacognitive knowledge (personal variables, task variables, and strategy variables). Having considered this classification of metacognitive regulatory skills in the existing research literature, I decided that the research was incomplete without the addition of audience inclusion. Due to the salient nature of communication of thoughts in essay writing, this strategy involved the part the audience played in reading the material. I therefore added the component 'considering the audience' to my classification of metacognitive regulatory skills.

While coding the data, I encountered several instances in which there appeared to be an overlap of codes in the same strategy used by a participant. In such cases, the statement was coded according to the strategy most strongly represented in the statement. Therefore, no single statement was coded as more than one strategy. In the example below, three strategies as italicised were detected:

(1)

"I think I just found a problem (*monitoring*) ... that I mingled different ideas, two sets of different ideas, in one single giant paragraph (*evaluating*), which would make readers lost (*considering the audience*)."

There were instances where the distinction between the strategies was unclear:

(2)

"I need to change the word 'arrange' (*monitoring*), because I should always avoid repetition of words (*strategy variable*)."

Below is another example of such overlaps, coded as three different strategies within the broad category of metacognitive knowledge:

(3)

"I know myself, I usually have difficulty (*person variable*) <u>expressing my thinking in proper English</u> (*task variable*), so what I can do is to outline my main ideas before I start writing (*strategy variable*)."

From the above examples, it is possible to comprehend how intricately convoluted are the interwoven strands of these strategies. These examples show the interconnectedness of

strategies and the fact that more than one can be present in a single utterance or thought episode. Attention to the subtle disparities of focus was thus imperative.

6.1.2. Results

The overall results of the protocol analysis are reported in Table 6.2 below. The frequency counts reflect participants' self-reported perceptions of the categories designed for this study. Altogether, a total of 1729 statements were coded from 24 think-alouds. It is interesting to note the differences between participants' think-aloud statements from session one to session four. The table presents information regarding the number of think-aloud statements for all participants. There were a total of 330 and 319 statements for participants +M+P (1) and (2) respectively, 293 statements for participant +M-P, 277 statements for participant -M+P, and 257 and 253 for participants -M-P (1) and (2) respectively. The highest and lowest totals are recorded in bold type. Accordingly, participant +M+P (1) generated the most think-aloud data, while participant -M-P (2) generated the least.

Overall, of the total number of statements coded, the lowest number was collected in the first session (n = 357) when participants were not yet completely familiar with the verbalisation activity, and the highest was obtained in the last session (n = 470) when participants were probably more accustomed to writing while thinking aloud. As Manchón, Murphy, and Roca de Larios (2005) and Manchón and Roca de Larios (2007) have pointed out, when participants complete several tasks, they gradually become more used to writing while thinking aloud, and thus their protocols become more informative in subsequent tasks.

Table 6.2 Summary of think-aloud protocols

Participant	Gender	Session 1 statements (minutes)	Session 2 statements (minutes)	Session 3 statements (minutes)	Session 4 statements (minutes)	Total statements (minutes)
+M+P (1)	Male	67 (51)	92 (49)	87 (53)	84 (50)	**330** (**203**)
+M+P (2)	Female	60 (49)	91 (50)	70 (52)	98 (53)	319 (204)
+M-P	Female	68 (55)	70 (53)	81 (57)	74 (58)	293 (223)
-M+P	Male	57 (56)	78 (50)	75 (53)	67 (57)	277 (216)
-M-P (1)	Male	52 (64)	71 (58)	62 (60)	72 (62)	257 (**244**)
-M-P (2)	Male	53 (59)	58 (56)	67 (58)	75 (61)	**253** (234)
Total		357 (334)	460 (316)	442 (333)	470 (341)	1729 (1324)

Note. M: Metacognitive instruction, P: Planning

Think-alouds by six individual participants in the course of four treatment sessions resulted in a total of 1324 minutes of recorded data. The mean time of writing the text aloud was 55.67 minutes (SD = 5.42) in session one, 52.67 minutes (SD = 3.67) in session two, 55.50 minutes (SD = 3.27) in session three, and 56.83 minutes (SD = 4.62) in session four. Having studied the literature on think-aloud protocols (e.g. Bannert & Mengelkamp, 2008; Ericsson & Simon, 1993; Janssen et al., 1996), I was aware that verbal reports could delay cognitive processing and would require participants to spend more time to perform the task. This was particularly evident for the no planning participants who seemed to have difficulties managing the time to complete the task. The two participants in the -M-P condition spent the longest among all (244 and 234 minutes for the first and second participant respectively). This was followed by the female participant in the +M-P condition (223 minutes) who similarly did not have the opportunity to plan prior to task performance.

Participants also differed in their ability to verbalise their ideas, particularly because this was taking place in their L2. In some parts of their verbalisations, participants were so engrossed in the writing task that at times they forgot to verbalise, and either remained silent while they were for instance rereading a paragraph they had just drafted or merely murmured in a low tone. These lapses could have been due to: a) limited capacity in short-term memory for performing the task and thinking aloud simultaneously, b) hearing the sound of one's own voice, and c) the influence of the researcher's verbal and non-verbal cues. Stratman and Hamp-Lyons (1994) have listed similar factors as possible causes of

problems. In those instances, I stimulated participants to think-aloud. Such prompts included "Continue thinking aloud", or "What are you thinking about at this moment?" Therefore these verbal protocols like any other are not a comprehensive report of metacognitive strategy use, but provide an indicative view of strategy use in action.

In the following two sections (6.1.2.1 and 6.1.2.2), the two components of metacognitive strategies detected in the think-aloud data are outlined, together with several examples and explanations. For the description of each strategy, the participants' strategy focus is shown in detail. The approach taken for the analysis is primarily descriptive, but also exploratory, where the nature of problems encountered by each participant, and the methods used to resolve the problems, varied from individual to individual. Manchón and Roca de Larios (2007) have postulated that the degree of dominance and the probability of occurrence of writing processes vary at certain stages as the writing task evolves (p. 579). Similarly, I speculated that given that the participants in the study received different instruction and were either provided with or deprived of the pre-task planning opportunities, they paid different amounts of attention to different processes at different stages of the composing process.

6.1.2.1. Metacognitive regulatory skills

There were instances where once the participants confronted a problem and contemplated on the reasons, they realised they could have made extensive use of a particular strategy, but did not. This shows that the participants did not recognise that the situation called for the use of that particular strategy. This finding supports Hartman and Sternberg's (1993) claim that learners may sometimes have the declarative (what information and strategies they have) and procedural knowledge (how to use them), but not the conditional knowledge (when and why to use them) required for application. Particularly when confronted with stress and competing fatigue and shortage of time, L2 learners need to be able to self-regulate their metacognitive knowledge and skills. After analysing the content of the data derived from the verbalisations, a number of statements that the participants reported in their writing practices were extracted. The four metacognitive regulatory skills which were operationally defined in this study are presented in the following pages.

Planning

Planning was the first code in the list of categories and the first strategy instructed in the first session. It is an important metacognitive strategy that can help learners to engage in preparation in relation to a learning goal. Moving on to a writing task, it is perceived that writers need to think about what they are required to accomplish and how they intend to go about accomplishing it. "Think-aloud protocols", Hayes and Nash (1996) acknowledged, "provide more detailed data about planning processes than any of the other methods" (p. 45).

Different views exist in the literature about how to operationalise, and thus code planning and formulation episodes. Some scholars (e.g. Cooper & Matsuhashi, 1983; M. Nystrand, 1982) do not distinguish between planning and formulation of text production in writing. They maintain that linearisation of language in sentences forms part of the planning operations and is not present only after the planning process has been finalised. Others (e.g. Hayes & Nash, 1996; Kellogg, 1994; Lansman, Smith, & Weber, 1990; Manchón & Roca de Larios, 2007) distinguish between the two because of the various strategies used by writers in planning prior to writing. The current study coded planning episodes based on Manchón and Roca de Larios' (2007, pp. 567-568) categorisation of planning corresponding to the following features:

1. The writer's position: the stance that the writer adopts including the topic, the audience, point of view.
2. The organisation of the text: decisions about how the ideas in the text are to be framed or sequenced in connection with the text type in hand and its rhetorical structure including introductions, conclusions, examples, paragraphs.
3. Ideational content: ideas generated by the writer that will be subsequently formulated in a linear fashion or discarded.
4. Procedural aspects: verbalisations about the procedures used throughout the composition such as the writer announcing in advance that s/he is going to write notes, reread, or introduce changes in the text.

Some sample segments of protocols that correspond to the four categories above are presented below:

(4)

"Both of these points of views have their pros and cons... there are those cases of marriages that in spite of having freely chosen their spouse turn up to be unsuccessful and there are those cases of arranged marriages that end up to be very successful. Now: which one do I favour? … And then how can I start something new… something surprising... to support that idea…?" (Participant +M+P (1))

(5)

"I can start writing about how young children nowadays are lack of social skills due to spending time alone or to study... I can't really come up with any statistical evidence or factual findings. So I might have to rely on my own personal experience or the people I know." (Participant +M+P (2))

(6)

"Now I should probably say that one of the positive sides of imposing strict discipline on children is to adapt them to a kind of ability to assimilate to the mainstream culture." (Participant +M-P)

(7)

"I'm re-reading this sentence but I'm not sure if it is very much relevant, so I put a star next to it and come back to it after I finish the whole paragraph." (Participant -M-P (2))

In the present study, all think-aloud participants initially glanced over the topic, task and instruction in order to comprehend the requirement of the task. Putting in this initial effort ensured that they were able to orient themselves more effectively towards the goals or sub-goals they accordingly specified. Based on this analysis, they formulated an action plan for the required task procedures. The planning variable was operationalised by first counting the total number of statements in the protocol which directly or indirectly referred to planning, followed by a comparison across participants. Planning statements not only involved the setting of goals in the form of written notes and outlines prior to starting the task, but these statements occurred at any point during the composition as long as they referred to those pre-linear decisions concerning (1) pragmatic aspects: the general stance the writer adopts (audience, writer's intentions, etc.); (2) discourse organisation: deciding

on the overall or partial structure of the text (text type, introductions, conclusions, points, examples, paragraphs, etc.); and (3) content: organising thoughts or ideas (Manchón et al., 2000, p. 19).

As shown in Table 6.3 below, participant +M+P (1) produced the most planning statements (n = 107) and participant -M-P (2) produced the fewest of all (n = 43). According to Sanguran (2005), the instructions that learners are given can influence, to a great extent, the way in which they plan. Participant +M+P (1) produced the most planning statements of all in the second session (n = 35), which might be due to the instruction of planning they received in the first session. This may suggest the transfer of strategy knowledge gained in one task to another. Also, participant -M+P produced almost as many statements as participant +M+P (2). Fewer planning statements for participant +M-P, compared to the two metacognitively instructed participants and even participant -M+P, can be explained by the absence of pre-task planning opportunity. Because this participant was required to plan online and write simultaneously, her attention may have been diverted from planning.

Table 6.3 Summary of statements coded as Planning

Participant	Session 1	Session 2	Session 3	Session 4	Total
+M+P (1)	20	35	27	25	107
+M+P (2)	17	24	21	22	84
+M-P	14	12	16	16	58
-M+P	15	21	26	21	83
-M-P (1)	12	15	13	17	57
-M-P (2)	10	7	14	13	43
Total	88	114	117	114	433

Note. M: Metacognitive instruction, P: Planning

According to studies conducted on the temporal configurations of writing strategies (e.g. Manchón & Roca de Larios, 2007; Rijlaarsdam & Van den Bergh, 1996; Tillema, van den Bergh, Rijlaarsdam, & Sanders, 2011; Van Den Bergh & Rijlaarsdam, 1996), the occurrence of metacognitive strategies varies not only at different moments during the writing process, but also between individual writers. In the most current study of the studies reported above, Tillema et al. (2011) observed that some activities such as planning are more likely to occur at the start of task execution, particularly for students

with higher planner scores compared with students with lower planner scores. Towards the end of task execution, they reported, the first group of students were less likely to apply planning activities than the lower-scoring students. Similarly, Manchón and Roca de Larios (2007) analysed the temporal distribution of planning throughout the composing process and ascertained that the allocation of planning time varies according to the type of writing task and the proficiency level of L2 writers. They empirically illustrated that the total composing time devoted to planning tends to be concentrated in the early stages of composition, with graphic outlines always appearing in the pre-writing stage.

Both participants in the +M+P condition made extensive use of graphic representations of their thoughts and knowledge in the pre-writing stage (e.g. via concept maps or semantic webs). They were not only instructed to draw graphic outlines, but were also given the opportunity to plan prior to approaching their writing tasks. They were taught the use of graphic organisers initiates with the generation of words relevant to the topic and the categorisation of them into groups. It then proceeds with the development of ideas on each topic and drawing linear or hierarchical organisational structures (Ojima, 2006, p. 568). This semantic mapping strategy has been widely implemented in a variety of instructional settings and has been reported to facilitate the process of L1 writing (e.g. Cronin, Sinatra, & Barkley, 1992; Pieronek, 1994; Washington, 1988) and L2 writing performance (Ojima, 2006; Schultz, 1991). Research conducted to examine how skilled and unskilled writers plan their ideas reveal that more skilled writers engage in global rather than local planning and unlike poor writers, they are better able to plan effectively, regardless of text content (Bereiter & Scardamalia, 1987).

Marzano, Pickering, and Pollock (2002) categorised graphic organisers into six common patterns: descriptive, time-sequence, process/cause-effect, episode, generalisation/principle and concept patterns. As can be seen in the following planning sheet (Figure 6.1) which belongs to participant +M+P (1)'s first-session writing practice, he first visually clustered the main topic of the writing task, i.e. arranged marriage, in the centre of the paper and lined it with the ideas associated with arguments for and against the central topic. He then grouped sub-ideas around this new topic. He also proceeded by a number of lexical searches and listed all the terms related although not all were later incorporated into his essay. The following diagrammatical demonstration corresponds to the concept pattern which, based on the Marzano et al.'s taxonomy, allows the learner to organise information

around a specific word or phrase that represents entire classes or categories of persons, places, things, and events. The characteristics or attributes of the concept, along with examples of each, are all included in this pattern. This representation presumably helped participant +M+P (1) to collect ideas more effectively and to understand the logic between ideas more adequately.

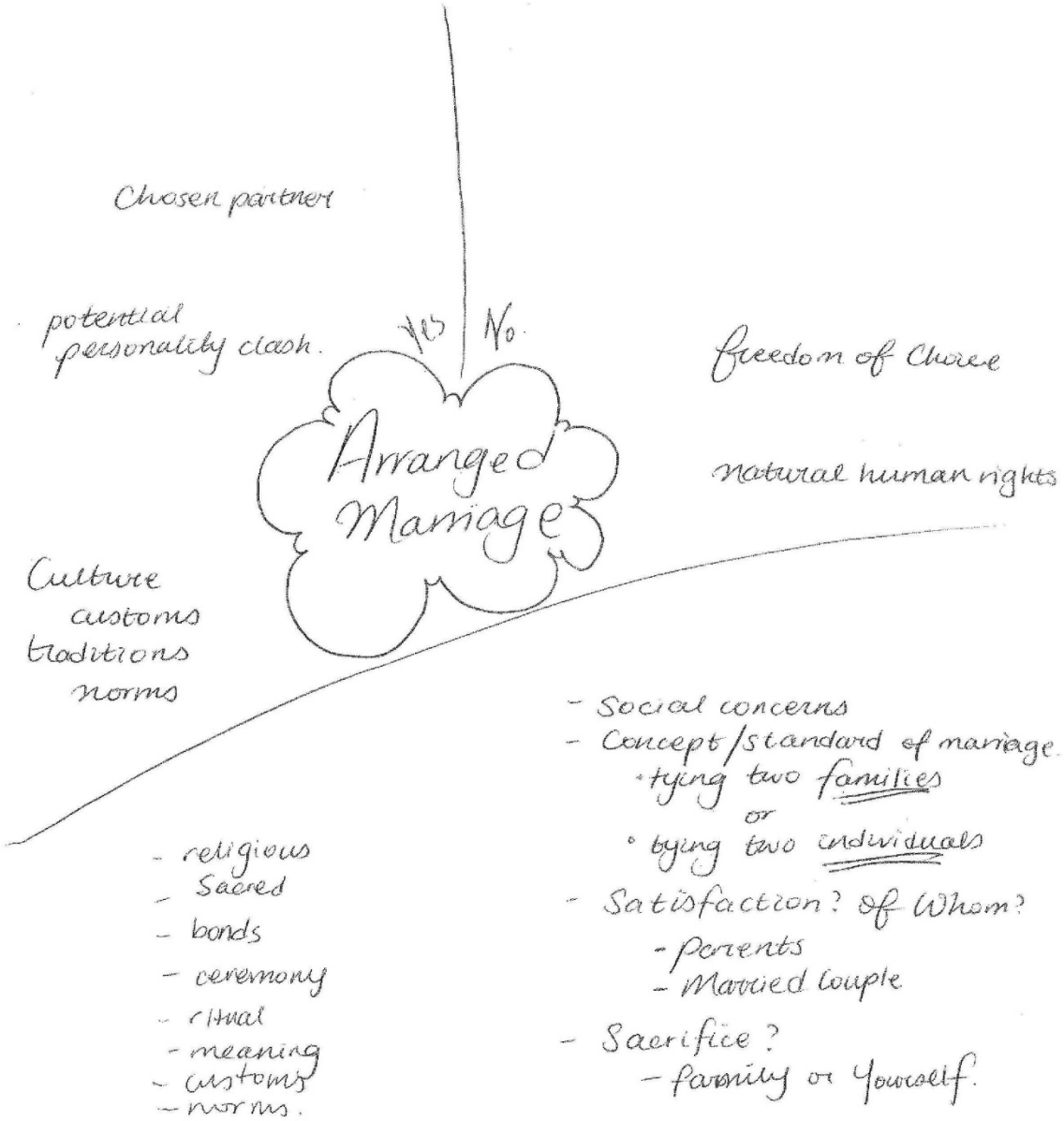


Figure 6.1 Participant +M+P (1)'s graphic organiser

Participant +M+P (2)'s planning sheet, as shown in Figure 6.2, has certain features in common with Marzano et al.'s (2002) descriptive pattern in the first half and

generalisation/principle pattern in the second half. The information organised into a descriptive outline does not need to be in any particular order while in a generalisation/principle pattern, information is arranged into general statements with supporting examples. The participant first underlined key words in the prompt and made a systematic approach to analyse and clarify the problem in order to ensure she completely understood the topic. In this way, she identified the purpose for writing and brainstormed ideas about the two systems of marriages to be compared. She recorded her thoughts below each two systems she had already placed on the planning sheet. She later decided which side of the argument she would support and recorded this information at the bottom of the planning sheet. She then used the notes generated earlier as a guide for writing. She continued to plan as she wrote and developed a text structure by adding details and examples in the introduction, body and conclusion paragraphs. She framed her pre-writing planning prior to actual composition in the shape of major points to be further developed. Such planned points functioned as guidelines during the execution stage. She interrupted her think-aloud at intervals to refer back to the points she had planned in advance. This type of planning equates what Cumming (1989) referred to as advanced planning, as opposed to emergent planning, where the planner engages in a process of reordering, addition, and substitution as their actual written texts progress. In this light, the participant structured her goals in a well-connected way and ensured she would stick to the top goals she had initially generated. Moragne e Silva (1989) interpreted this goal-to-text approach to planning as an evidence of skilled writing performance. This approach, she claimed, could help the writer to initially generate ideas, then generate goals, and finally to transcribe the text.

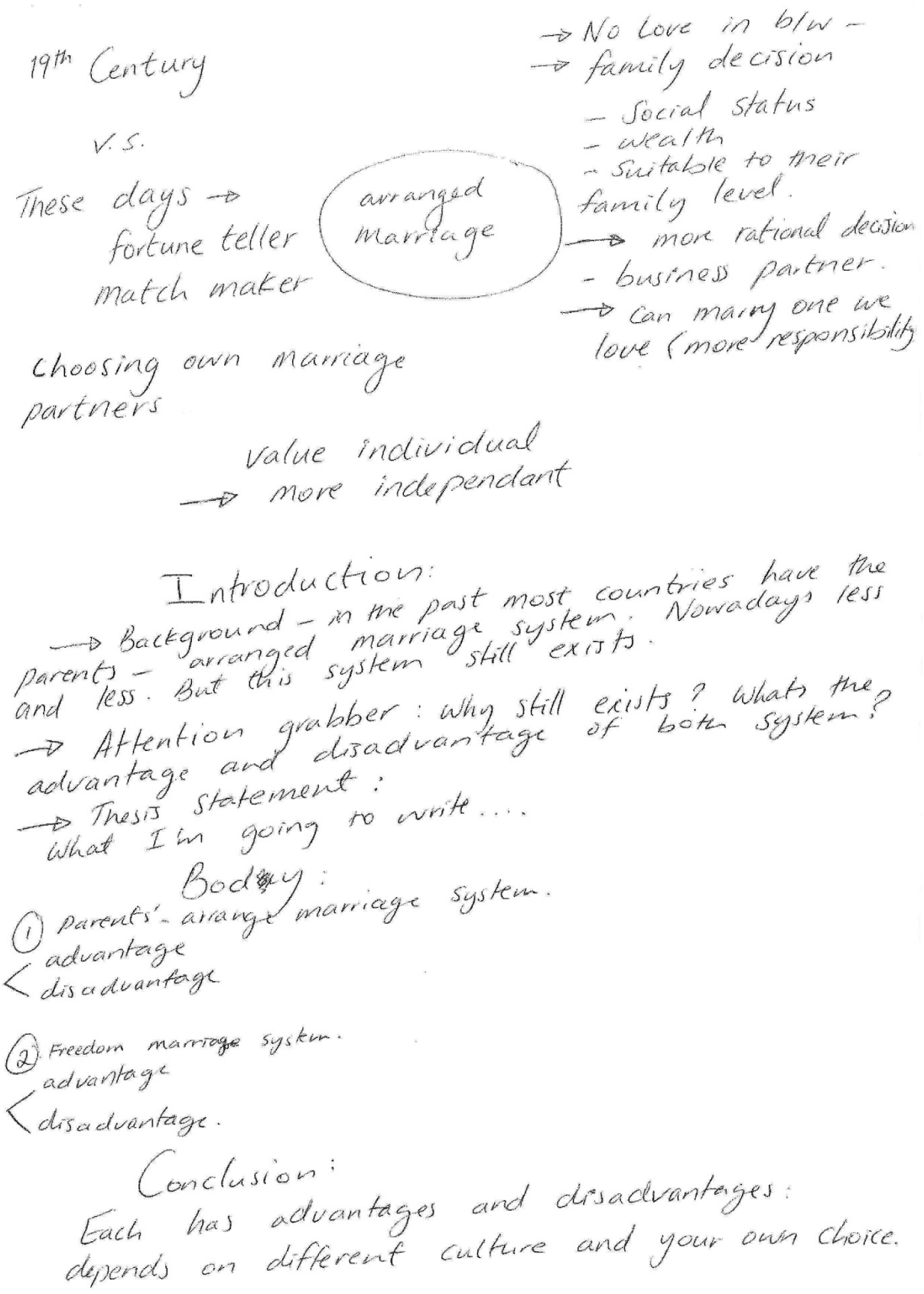


Figure 6.2 Participant +M+P (2)'s graphic organiser

From prior research, it was established that even when explicitly prompted to participate in advanced planning, struggling writers devoted less than half a minute to planning, regardless of their age, the writing genre, or the writing medium, e.g. handwriting, typing, or dictating (De La Paz, 1999; Graham, 1990; Harris et al., 2006; Lane et al., 2008). In the present study, although participant -M+P did not receive metacognitive instruction on how to plan and what to plan for, he was prompted to take advantage of his 10-minute pre-task planning opportunity. Yet, at the beginning of his first writing practice he stated his lack of interest in the need to plan,

(8)

"I do not feel I need to plan. I can write about the topic just like I would talk about it in an oral discussion. Due to the time restriction, I probably would have skipped the planning process, but because I should think aloud, I am required to do planning."

His planning sheet in the first session corresponds to Marzano et al.'s (2002) generalisation/principle pattern. As shown in Figure 6.3, he illustrated the way he brainstormed to generate ideas based on his pre-planned division of the essay into three paragraphs of an introduction, body, and conclusion. He started with a question in the introduction and emphasised it to be stated by two different opinions. In the body paragraph, he jotted down three possible arguments but did not seem worried about whether and which of these ideas were to be used. More than anything else, this type of organisation served as pointers for the content retrieval during the composition phase. The analysis of the actual composition showed that the paragraphs closely followed the order of retrieval of ideas which were thought of during the planning phase. The note for a conclusion paragraph was left without any comment. Although he was not under time pressure, he stated that he hoped he would have more time through the composing process to decide upon the tension between the two sources of decision making (parents vs. individuals) and to express his own views as to which one would more successfully contribute to happiness in marriage. In the actual writing text, a conclusion paragraph was produced as he proceeded with the actual text.

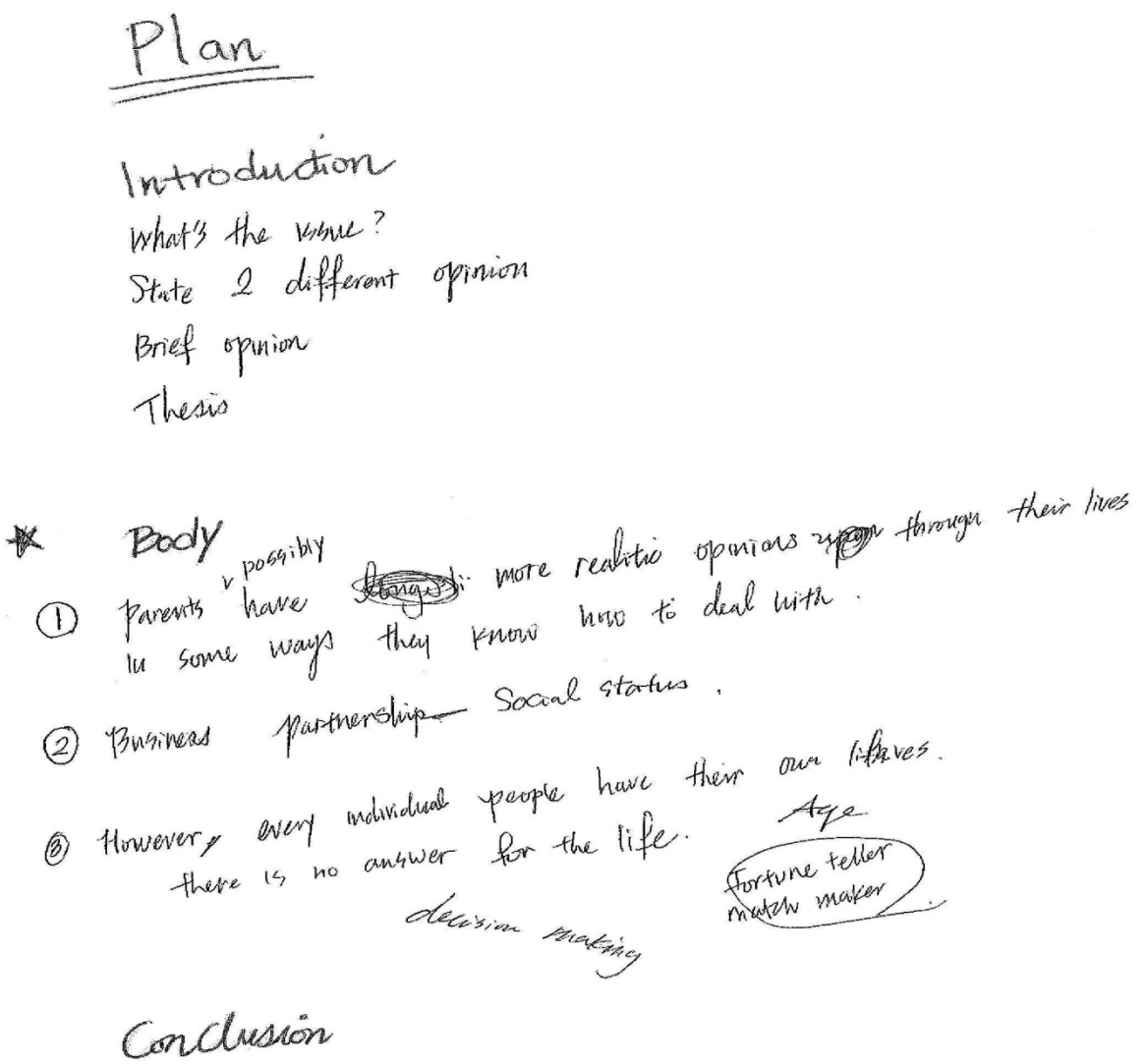


Figure 6.3 Participant -M+P's graphic organiser

As can be seen from these three graphic organisers, the provision of planning enabled the participants to manoeuvre through the planning sheets to translate abstract information into a more concrete written form. The planning activities mainly consisted of setting the goal, managing how to accomplish the task, organising the structure and content sequence, and generating key phrases accordingly These written planning notes and outlines prior to the text production provide credible external evidence for participants' planning processes, although as Hayes and Nash (1996) have pointed out these notes may be more sensitive to some kinds of abstract planning, such as organising and brainstorming than to language planning, which results in text, but usually not in a written plan (p. 46). The differences in participants' use of graphic organisers can mainly be attributed to their familiarity with the

planning activities, and the type of instruction they received. The two participants in the metacognitive groups not only spent longer time to plan than the participant in the non-metacognitive groups, but also, as shown in their sample graphic organisers, planned a more detailed overall organisation. Similarly, in her study on the impact of concept mapping as a pre-task planning activity, Ojima (2006) claimed that various factors seem to affect the learners' decision-making process in using or not using concept mapping. These factors include learners' familiarity with the strategy, the nature of writing tasks (e.g. length, complexity, discourse types), and situational restrictions (e.g. time limits). A sense of benefit that learners expect to obtain from planning activities, Ojima believes, can also affect their motivation to engage in writing tasks (p. 582).

Although Van Den Bergh and Rijlaarsdam (2001) have observed that planning activities are most effective during the initial stages of writing process, the think-aloud protocols showed that planning does not necessarily occur prior to the essay composition in the form of pre-task planning. According to Manchón and Roca de Larios (2007, p. 556), planning can occur both globally during the text creation process and at different stages of the composing process (i.e., at the beginning, middle, and end). Yet, they observed that planning episodes mainly occurred in the first third of the composition process. As Hayes and Flower (1980) postulated, any sub-processes of writing could interrupt or integrate with any others. Thus planning may be evoked at any cognitive processes when the writer devotes the time to plan at the beginning, middle, and even end stages of the composing process. This emphasises the recursive nature of the writing process. Cumming (1989) named those who created plans early in the course of composing as advanced planners, and those who planned during the course of writing as emergent planners.

As can be seen in the following examples, when planning the writing task, participant - M+P, who had no metacognitive instruction, sometimes involved himself in the strategy of self-questioning. It seemed to take him a greater length of time, however, to decide upon what should follow the question he raised:

(9)

"How can I start something new… something surprising... to support that idea...?"

As a result of his steering capacity to monitor the usefulness of his planning efforts and to (re)formulate goals which the text needed to satisfy, he constantly asked himself if his

strategy was working and attempted to change it if not. The same participant had shown some sort of confusion and disorientation at the outset of the writing process:

(10)

"I have no idea how to begin this writing…"

These two examples show that this participant lacked the metacognitive ability to take time to prepare and plan what needed to be accomplished, and once he stopped to monitor the flow of writing, he could not recognise how to incorporate the use of proper writing strategies. This was in contrast to the metacognitively instructed participants who were taught how to prepare and plan, how to set appropriate writing goals, and how to begin a writing task.

There were also several instances where planning was coded as a strategy when a participant was involved in what can be described as 'conditional planning' while monitoring writing, for instance, the body paragraph:

(11)

"If I need some more supportive ideas for my opinion, I may insert a third paragraph later."

This indeed facilitated participant +M-P's reflection upon her writing, as she began to notice possible unplanned problems. On the one hand, she probably became aware of distracting or intruding stimuli. On the other hand, she planned or adjusted how to efficiently cope with those and sustain effort over the total time of the task. These skills are in fact metacognitive in nature, because they involve executive functions. At some latter stages in the second writing practice, she stated:

(12)

"If I could plan the general structure and some very small details that needed to be included in each paragraph, my composition would come out better."

In the third writing practice, she similarly stated:

(13)

"I did not know much about the topic. If I could make an outline before starting to write, I could probably produce more organised ideas."

This reflects Manchón and Roca de Larios (2007) study which confirmed how the temporal nature of planning depends on prior topic knowledge and topic familiarity in a context where the participants were investigated across languages and three different levels of proficiency. They observed that when writing about a less familiar topic, the expert or higher proficiency writers not only spent more time to construct their pragmatic, textual, and ideational representations before the composition stage, but also reflected more on the organisation of their ideas. These findings were previously reported in a similar way in L1 writing by Olive, Piolat, and Roussey (1997), and Piolat, Roussey, and Rous (1996). They had found that undergraduate students who were skilled writers in L1 planned more extensively when writing about a topic they knew little about, whereas unskilled writers planned less in a similar situation.

Participant -M-P (2) generally plunged into writing the first sentence right after he read the prompt and wondered what he could write about the topic. These first sentences were mainly copied from the prompts. At times, he expressed some levels of disorientation and voiced dissatisfaction with his ability to develop and conclude an argument:

(14)

"I am lost. This argument is not working. I can't get anywhere like this. I shouldn't have started it at all".

This confusion might have been resolved if he had been provided with the opportunity to plan prior to the task. Not having sketched a pre-planned set of arguments created a challenge for him and hampered the progress of his composition.

Considering the Audience

I added considering the audience to the list of strategies, because of the significance of recognising writing as a task performance which constitutes social as well as cognitive activities. It was my perception that knowing who the writing is being directed towards defines the form and style which the writing eventually takes. It is pertinent to add at this point that my decision was endorsed by Bereiter and Scardamalia's (1987) findings that good writers were better able than poor writers to adopt readers' perspectives. As illustrated in Table 6.4, during the first writing practice, the lowest number of participants reported having considered the audience. In total, participant +M+P (2) reported the most statements (n = 16), and participant -M+P reported the fewest (n = 9).

Table 6.4 Summary of statements coded as Considering the Audience

Participant	Session 1	Session 2	Session 3	Session 4	Total
+M+P (1)	3	5	2	2	12
+M+P (2)	3	5	4	4	16
+M-P	1	4	3	5	13
-M+P	0	2	4	3	9
-M-P (1)	1	1	5	3	10
-M-P (2)	2	3	3	3	11
Total	10	20	21	20	71

Note. M: Metacognitive instruction, P: Planning

More importantly, the differences among groups lay in the fact that in the second session three participants in the metacognitively instructed conditions were explicitly instructed to create an audience. They tried to picture one that was distinguishable and then aimed to write persuasively for that typically perceived audience, as in the following example extracted from participant +M+P (1)'s second session protocol data:

(15)

"If I can show my experience so enthusiastically, my audience – the school children parents – will get interested in my topic, and then I have been successful, because what they really want to hear are real stories."

Participant +M+P (2) showed an awareness of addressing familiar audience vs. audience of strangers in the second session:

(16)

"I write differently when I write to a friend than when I should write to a stranger."

Or as in the example below, participant +M-P, who was taught that an effective argumentative writing style involves thinking about the audience and its persuasive purpose in writing, consciously monitored the use of this strategy and paused while writing in order to ask herself about the clarity and intelligibility of the essay:

(17)

"…I still wonder if this viewpoint makes sense to the audience… Maybe I should make this idea more clear to read."

This understanding of audience appeared to maximise her potential to reflect upon the appropriateness of the language and to provide the right amount of background information for the intended audience. This way, she might have also thought to make a more effective use of examples in supporting her arguments.

Participant -M+P, however, not only ranked the lowest in the frequency of stating the consideration of audience (n = 9), but also a detailed analysis of his think-aloud data showed that he merely thought whether his essay could clearly state what he intended to express. The example below demonstrates that the general attention to the level of formality was at times taken into consideration, without a particular sensitivity to an intended audience in mind:

(18)

"But this doesn't sound very formal, I mean, a very academic essay."

Similar patterns of implicitly considering the audience were witnessed in other participants who were not explicitly instructed to attend to the needs of an intended audience. Overall, while none of these three participants explicitly vocalised any degree of awareness of this strategy, particularly in the first session, they all implicitly noted statements which enabled me to apply the appropriate coding.

It can, therefore, be expected that participants who were taught to consider an audience in their writing would be more likely to have a conscious awareness of its existence and thus would make better use of this strategy while writing. This explicit instruction also reflects Sengupta's (2000) study, in which the concept and importance of reader appeared to gradually take shape in students' minds after an explicit training of revision strategies to a group of female secondary-school students in Hong Kong. After all, if learners do not know of a strategy, they might not be able to use it (Pintrich, 2002). This has led me to conclude that metacognitive instruction can play an important role in the accomplishment of the task and, by implication, in learners' performance.

Monitoring

The third metacognitive strategy reported in this section is Monitoring. This term here refers to continually checking the progress, or lack thereof, on the task at hand and making judgments about the strength and weakness of the performance. In this study, all

participants provided several examples indicating the application of monitoring strategy, which is why this strategy marked the most number of statements in terms of the relative frequency of verbalisations in the data (n = 579). The total number of utterances that signified monitoring strategy was the most for participant +M-P (n = 106), and the fewest for participant -M+P (n = 88). The most monitoring occurred in the third session for participant +M-P (n = 37). Interestingly, this was the session when the metacognitive groups were instructed in how to monitor their performance.

Table 6.5 Summary of statements coded as Monitoring

Participant	Session 1	Session 2	Session 3	Session 4	Total
+M+P (1)	20	23	25	25	93
+M+P (2)	17	30	24	28	99
+M-P	21	25	37	23	106
-M+P	17	26	25	20	88
-M-P (1)	19	27	21	25	92
-M-P (2)	22	25	20	34	101
Total	116	156	152	155	579

Note. M: Metacognitive instruction, P: Planning

In the present study, there existed some comments in participants' verbalisations which represented their recognition of a problem but signified that they did not attempt to or lacked the ability to correct or improve the texts. Alternatively, there were some other comments which characterised the presence of some higher-level evaluative capacity to revise the texts. In a study of EFL Spanish learners, Roca de Larios et al. (1999) used think-aloud method to explore learners' alternative strategies when their original syntactic plan was not satisfactory for a variety of linguistic, ideational or textual reasons. These researchers traced the patterns the learners used in their strategy use to compensate for the lack of linguistic resources typical of L2 learners. They concluded that L2 writers used restructuring strategies when they lacked access to relevant linguistic knowledge and thus were unable to express their intended meaning in a foreign language.

The following statement reinforced participant +M+P (2) monitoring the time spent on task. As the text was advancing steadily, she paused with the following breakdown in the composing process:

(19)

"I am probably spending too much time on this particular view. I should probably start another one. Let me try the opposing view now".

This echoes Manchón and Roca de Larios' (2007) assertion that writers monitor the composing processes in the course of text production and decide which processes to stop and which to activate at a precise moment (p. 551).

It is important to note that monitoring may occur at linguistic and discourse levels. Out of the total utterances within this category, many concentrated on surface form: 52.68%, 54.54% for participants +M+P (1) and +M+P (2), 58.49%, and 63.63% for participants +M-P, -M+P, and 51.08%, 64.35% for participants -M-P (1), and -M-P (2). These monitoring instances generally occurred during the writing of a sentence, rather than between sentences or while reading over a larger chunk. Below, two examples clearly demonstrate the degree to which these two levels can work hand in hand. In the first example, while formulating her text, participant +M-P was solving a linguistic problem, which was lexical in nature. She was closely regulating the lexical appropriateness of the verb hit. She decided to change it to the verb punish in order to eliminate the negative association the audience might make with the physical punishment:

(20)

"My parents hit me when I was a kid... um.. I think if I say 'hit', that might be a little weird, I mean too strong, I'd better say 'punished'.

In the second example, participant +M+P (2) was monitoring the task at the discourse level:

(21)

"This idea doesn't really link to the previous paragraph. Maybe I can change it to a better one… Yes, I should."

As a result of her concern for evaluating whether her writing made sense, she worked out the actual direction of her argumentation quickly and accurately.

From these two examples, it can be observed that the former is at a linguistic level, whereas the latter is at a discoursal level of monitoring. This dynamic nature of monitoring

while formulating echoes Raimes's (1985) interpretation of non-linear composing process. She argued that "writers inevitably discover new ideas as they write and then change their plans and goals accordingly" (p. 230). As shown in the second example above, monitoring helped the participant to read back over phrases and sentences just written, then to analyse the expressed idea, and finally to plan what subsequently needed to be written.

Sometimes monitoring could assist the participants to backtrack in order to either generate more text or check back whether a written statement successfully matched with its intended meaning. In doing so, participants who had the pre-task planning opportunity from time to time rescanned their written notes or outlines. This demonstrated the recursive nature of the writing process in which the writers move back and forth between the sub-processes of planning, transcription and revision (Manchón et al., 2000). At times, participants paused to judge how well or badly they were writing, that is, whether they were implementing the strategies they had selected and whether these strategies were helping them to achieve the goal as intended. In doing so, hesitations and self-questioning strategies were frequently used in the monitoring phase of the composition.

(22)

"I'm not sure if this is the right way to describe the status in my country, is it? Have I missed more important things?" (Participant +M+P (1))

Or participant +M-P paused to evaluate the idea she had generated, but decided to restructure the argument which was going astray.

(23)

"I have written about marriages in different countries… No, this does not seem relevant… Let me see… No, I should concentrate on my home country.. but maybe in different generations?"

According to Hartman (2001), a self-questioning strategy can help learners to improve self-awareness and a control over their thinking and to enhance their performance before, during, and after their task completion (p. 55). Hesitation was also observed in participant -M+P who did not receive the explicit metacognitive instruction:

(24)

"But I don't like this way of comparing these two views… Well, maybe I am wrong."

As can be seen, knowing when to stop and revise one's writing does not seem an easy task, but knowing how to edit it when faced with challenges is clearly a more difficult task. There were many instances where participants could see a linguistic or discourse problem in their writing, but had more difficulty fixing that problem. Study findings support Delclos and Harrington's (1991) suggestion that training and practice can help improve monitoring ability.

Evaluating

Evaluating here relates to actively controlling over the judgments one makes about the strengths or weaknesses of linguistic, textual, ideational and pragmatic choices in the task accomplishment and in accordingly changing or modifying the task. After a writing task is completed, the writer needs to evaluate and judge whether the written outcome is cohesive at a linguistic level, coherent at a discoursal level, and relevant with respect to reaching the specified goals at the planning stage. This is one of the crucial areas where the difference between good and poor writers arises. With regard to text revisions, Bereiter and Scardamalia (1987) found that experienced writers had less difficulty than poor writers diagnosing text problems and correcting them.

In this study, as illustrated in Table 6.6, the total number of statements coded as evaluating strategy was highest for participant +M+P (1) with 43 statements and the fewest for participant -M-P (1) with 21 statements. The highest frequency of monitoring occurred for participant +M+P (2) in the fourth session (n = 17). The lowest frequency of monitoring was for participant -M-P (1) in the first session (n = 3). The comparatively less frequent episodes of evaluating compared to monitoring in all participants' think-alouds were most probably the result of fewer reports or statements in the summary of my observations on this strategy.

Table 6.6 Summary of statements coded as Evaluating

Participant	Session 1	Session 2	Session 3	Session 4	Total
+M+P (1)	7	9	15	12	43
+M+P (2)	5	5	8	17	35
+M-P	11	13	8	9	41
-M+P	7	7	6	9	29
-M-P (1)	3	7	6	5	21
-M-P (2)	7	6	8	10	31
Total	40	47	51	62	200

Note. M: Metacognitive instruction, P: Planning

Research suggests that revising, like planning, as a critical self-regulatory process in skilful writing is what distinguishes skilled writers from novice and struggling writers (Graham, 2006; Graham & Harris, 2000; Harris et al., 2010; McCutchen, 2006; Sitko, 1998). While skilled writers attend to both linguistic and conceptual aspects of their texts to increase their overall quality of their compositions, the revising behaviours of novice and struggling writers demonstrate that they focus their attention nearly exclusively on the surface-level textual features of discrete words and sentences, rather than the macro-structure and meaning of their compositions (Harris et al., 2010; McCutchen, 2006). MacArthur and Graham (1987) reported that around 60% of fourth-grade and fifth-grade struggling writers' revisions targeted spelling, punctuation, capitalisation, or handwriting. In the present study, out of the total utterances coded as monitoring strategy, some concentrated on surface form: 53.48%, 48.57% for participants +M+P (1) and +M+P (2), 58.53%, and 55.17% for participants +M-P, -M+P, and 57.14%, 61.29% for participants -M-P (1), and -M-P (2).

It was found that, as Ku and Ho (2010) have also posited, there was a relationship between monitoring and evaluating strategies. Following participant +M+P (1)'s checking and validating the clarity of the information produced as his supporting sentences, he reassessed the argument and the weight he had placed on the claim he had made in his text:

(25)

"What I have written does not state how generations have changed over time… and in what respect. Yes, I should include a couple of sentences at the start of this

paragraph showing the historical background, before arguing which one is more agreeable nowadays."

In contrast, the following comment illustrates that -M-P (2) identified the inconclusiveness of his reasoning and thought product, which is evidently metacognitive in nature, but he took no actions to re-evaluate the task progress:

(26)

"Or is it my general belief that marriages arranged by parents are quite outdated and have no place in modern age? … But there may be some marriages who still favour arranged types… Maybe I should have also included the advantages… But it will take so long, so maybe not."

This relationship between monitoring and evaluating reinforces that "effective monitoring depends on whether monitoring activities leads to a re-examination of how one should approach the task (i.e., evaluating strategies)" (ibid. p. 262).

There were instances in the protocols that revealed whether participants had concentrated their efforts on surface-level features of their written texts or whether they had attempted to improve the conceptual aspects or global structure. Once participant +M-P finished writing her essay, she rescanned her essay and attempted to relocate a sentence from the introductory paragraph to the concluding paragraph:

(27)

"I'd better cross out this sentence in the introduction and put it in the conclusion paragraph."

This example demonstrates that she had evaluated what she had written and modified a point already dealt with. It appears that this conscious application of thought and decision making benefited the intended thrust and direction of a written output, suggesting the writer's emerging reader awareness.

Another factor that played a crucial role in the extent to which participants engaged in evaluating their written outcome was comparing the text against the rubric of the writing task. Once participant +M+P (2) finished writing, she referred back to the rubric of the

prompt to evaluate whether she had comprehended it well enough or whether she had achieved the overall writing goal:

(28)

"Did I comprehend the topic? Let me check if I covered all the arguments."

This shows that she evaluated the writing outcome, with respect to the overall writing goals.

When participant -M+P completed his composition, he rescanned segments of his essay and added another opposing idea which he had neither jotted down in his pre-task planned outline, nor during the composition. This aptly reflects how re-evaluating a text "may reactivate in the writer's head the rich pool of unwritten possibilities which may lend themselves to being used as input for subsequent transcription processes" (Roca de Larios, 1996, p. 201).

6.1.2.2. Metacognitive knowledge

In general, the metacognitive regulatory skills, described in section 6.1.2.1, were not used in a linear way to move from preparing and planning, to monitoring, to evaluating the task. The think-aloud data suggest that different processes might be occurring at the same time during the writing task. This also applies to the second component of metacognition which mainly consists of knowledge of general strategies, and not the actual use of those strategies. This is, in fact, the extent to which the knowledge of metacognition can help learners to become knowledgeable of and responsible for their own cognition and thinking. After all, once writers are aware of their strengths and weaknesses (knowledge of person), the nature of the task at hand (knowledge of task), and the available tools or skills (knowledge of strategy), they can perform the task more effectively. Out of the knowledge areas that Hill and Hannofin (1997) examined, metacognitive knowledge had the most influence on strategy use, and participants with high metacognitive knowledge were better able to reflect on their processes and refine their actions. They were also better oriented to consistently monitor their thinking and revise their action accordingly (p. 56). The three areas of metacognitive knowledge are elaborated below.

Knowledge of Person

In terms of writing, self-knowledge can be a significant facilitator. Learners who are aware of their own strengths and weaknesses in writing can adjust their writing to the range of diverse problems that may arise during the process of composition. The summary of statements coded as knowledge of person, as shown in Table 6.7, demonstrates that participants made statements with almost equal frequency about their perceived selves as writers.

Table 6.7 Summary of statements coded as Knowledge of Person

Participant	Session 1	Session 2	Session 3	Session 4	Total
+M+P (1)	5	3	3	4	15
+M+P (2)	2	5	5	4	16
+M-P	3	3	6	5	17
-M+P	7	5	3	3	18
-M-P (1)	2	4	6	3	15
-M-P (2)	4	3	5	4	16
Total	23	23	28	23	97

Note. M: Metacognitive instruction, P: Planning

As Harris et al. (2009, p. 134) have pointed out, this knowledge of self as a writer can include forms of writing in which the writer has been successfully or unsuccessfully engaged in the past, components or elements of writing they are comfortable with or have not yet mastered, and the environmental characteristics they prefer. For example, participant +M-P realised that she did not know much about the marriage procedure in other countries, the topic of the first session writing task. Undeterred, she drew on personal background knowledge of the circumstances in her own country.

(29)

"I don't know much about the customs and traditions in other countries, but even in our country, there are different social classes in different parts of the country which follow different marriage rituals."

Comments such as these on one's ability/inability or knowledge/lack of knowledge reflect that, as Schraw (2001) has rightly argued, metacognitive knowledge can at times

compensate for lack of relevant prior knowledge (p. 7). In this respect, drawing upon a wealth of topical knowledge obtained through experience and exposure in a specific context, this participant used several examples to make sure she covered the topic thoroughly.

Similarly, when participant -M+P realized that he had difficulty writing the minimum number of words required for the essay writing (250 words), he attempted to prepare for the upcoming writing task in an appropriate manner by allocating and adhering to fixed time limits for particular task components. However, if learners do not have knowledge of their own strengths and weaknesses, they will be less likely to adapt themselves to different situations. For instance, if participants read the topic of the writing and think they understand it, but in reality they do not, they will be less likely to go back and reread or review the whole composition to make sure it is understood, because they have already run out of time. Yet if they have an accurate perception of their lack of knowledge, they will be able to regulate the task and maintain a pace to conclude the task in the time given. For instance, participant –M+P stated:

(30)

"I sometimes lose marks in writing because I misunderstand the topic."

Another aspect of metacognitive knowledge of self is one's perceived self-efficacy which refers to "a personal judgment of one's capability to execute actions required to perform" (Hill & Hannofin, 1997, p. 39). Self-efficacy is particularly determinant when learners need to deal with stressful situations. Harris, Graham, and Pressley (1992) indicated that many struggling writers' self-efficacy for writing did not match with their actual performance. In effect, they overestimated or underestimated their abilities and motivation to write. In line with this, the following statement reflects participant -M-P (2)'s uncertainty as to how to proceed to the conclusion paragraph when he was not confident enough about the arguments expressed in the body paragraph:

(31)

"I am looking at the body paragraph and it doesn't make sense to me. I am not sure why it turned out to be like this. Maybe I wasn't careful about the choice of my arguments. I'm stuck. How can I wrap this up?"

This resulted in a defining moment when it was necessary for this participant to decide how much time and effort he required to effectively achieve the desired result.

Knowledge of Task

Throughout the process of writing, individuals accumulate a variety of knowledge about different cognitive tasks. They also have some prior knowledge and experience related to mechanics, forms or styles in different types of essay writing, the kinds of processing required and the information-processing demands a task might place on the writer. This knowledge of task is a principal factor in influencing the learning and retention of the new material. The analysis of the think-aloud protocols in this study, as shown in Table 6.8, indicated that participant +M+P (2) stated the most expressing the knowledge of task (n = 39), whereas participant -M+P expressed the least of all (n = 21).

Table 6.8 Summary of statements coded as Knowledge of Task

Participant	Session 1	Session 2	Session 3	Session 4	Total
+M+P (1)	8	10	7	11	36
+M+P (2)	6	17	4	12	39
+M-P	10	10	5	6	31
-M+P	4	5	5	7	21
-M-P (1)	5	6	8	10	29
-M-P (2)	6	9	11	6	32
Total	39	57	40	52	188

Note. M: Metacognitive instruction, P: Planning

In the following example, knowledge of task for participant +M-P possibly included the knowledge that argumentative writing tasks are more difficult than other modes of writing and thus may require different cognitive strategies:

(32)

"This is an argumentative writing (task variable). As it was instructed by our teacher, I need to write my views in at least four paragraphs; one introduction, two body paragraphs, but with two different views, and one conclusion."

This knowledge of task nature reflects the what and the how of different strategies and if this formulaic and declarative knowledge is never shared through discussion or explicit instruction, it is difficult for learners to obtain. The above transcript also illustrates how the participant attempted to apply the instructor's recommendations rigidly.

Knowledge of Strategy

In addition to the need for knowledge of self as a writer and the writing task, knowledge is required regarding the strategies or procedures available for achieving various cognitive goals, as well as knowledge about the relative effectiveness of available strategies for achieving these goals. This knowledge of strategies may be general, such as writing an attention-grabbing thesis statement or may be specific to a particular writing task, such as writing a narrative or an argumentative essay. Among all, participant -M-P (1) produced the most knowledge of strategy statements in his think-alouds (n = 33), while -M-P (2) produced the fewest (n = 18).

Table 6.9 Summary of statements coded as Knowledge of Strategy

Participant	Session 1	Session 2	Session 3	Session 4	Total
+M+P (1)	4	7	8	5	24
+M+P (2)	10	5	4	11	30
+M-P	8	3	6	10	27
-M+P	7	12	6	4	29
-M-P (1)	10	11	3	9	33
-M-P (2)	2	5	6	5	18
Total	41	43	33	44	161

Note. M: Metacognitive instruction, P: Planning

In the following example, participant +M+P (1) modified his use of strategy based on the awareness of the effectiveness of what he calls 'rereading' chunks of the text in order to establish the course of his argument:

(33)

"I'm reading again because reading can make me think more and develop my ideas. Because by reading, by rereading my essay, I usually get some ideas and I can…even if I go off… go away from my thesis statement and my intention, I can get back on track to my way, original way, so sometimes I need to reread."

Thus knowledge of rereading strategy helped him to resolve potential ideational problems which could arise if he were not aware of this strategy.

Summary

The aim of protocol analysis in this project was to examine whether and to what extent participants remembered and relied on the strategies that were taught to them during the treatment. This also helped investigate whether they transferred the strategies they had learned during each treatment session to further analogous writing tasks. In addition, it helped to explore whether those who did not have the metacognitive strategy training were still making subconscious use of the strategies or not, as it has been argued that individual learners bring their own skills, experiences and strategies to tasks (Pintrich, 2002). Using the think-aloud method was time-consuming and labour-intensive in both administration and processing of the raw data. Therefore, the current study collected and analysed the data of only six participants. In fact, having more than one volunteer to think aloud helped monitor participants' unique engagement in the planning and composing activities although they had received different instructions. The results of protocol analysis showed variation in the focus and frequency of utterances during think-alouds. It is important to note that variation can both oscillate between individuals, and, for any one individual, from one session/task to another. More importantly, the results also revealed that the number of times a participant's statement was coded as a strategy could not be taken as the only indicator of success or failure. Instead, the distinguishing factor was clear patterns of differences in the use of strategies. The findings indicated that learning strategies, person and task characteristics were relevant when mastering a specific situation. Moreover, writers needed to identify, plan, and monitor their writing needs, which are all fundamental strategies in the successful completion of a task. According to Hill and Hannofin (1997), "weak metacognitive knowledge and skill may limit learners in defining learning needs, evaluating available resources, and revising their learning strategies, while strong metacognitive knowledge and skill are likely to improve learning" (p. 38).

Furthermore, the protocol data demonstrated that participants were able to articulate metacognitive regulatory strategies and metacognitive knowledge while speaking about all the seven types of strategies. The think-aloud protocols provided an important source of data on learners' on-line use of strategies, and served to complement the larger data set of this study. There were no attempts, however, to generalise the results to a broader

population. As Schellings et al. (2012) have pointed out, the think-aloud method is a valuable metacognitive assessment method for tapping conscious reflections only, and thus the verbal reports of participants' thought sequences provide an approximation of what actually takes place. In spite of a small-scale examination, the present study provided significant insights on identifying participants' concurrent thinking processes along with thinking performance. N. J. Anderson (2007) indicates that "metacognition is not simply thinking back on an event, describing what happened and how one felt about it. It requires a cognitive awareness and engagement with the awareness of one's thinking" (p. 170). He encourages teachers to engage students in think-aloud protocols in writing classrooms to make their thinking more visible while writing. In the present study, think-aloud method was utilised in order that the participants could consciously articulate the knowledge about what and how they write. Flower and Hayes (1981) maintain that "part of the drama of writing is seeing how writers juggle and integrate the multiple constraints of their knowledge, their plans, and their text into the production of each new sentence" (p. 371). Despite the fact that the results of this analysis provided useful input for the general discussion on how learners apply the training of metacognition during the writing process, further empirical investigations were carried out via questionnaires to explore the effectiveness of metacognitive knowledge and the utility of metacognitive regulatory skills in writing performance.

6.2. Metacognitive Strategy Questionnaire (MSQ)

Questionnaires are widely used methods for researching writing, and are particularly useful for "collecting large amounts of structured, often numerical, easily analysable self-report data" (Hyland, 2010, pp. 195-196). Messick (1995) describes two main threats that may compromise the construct validity of measurement methods in education and psychology. The first one is construct-irrelevant variance, by which he means that the measurement may be too broad and may contain some irrelevant information. The second threat is construct-underrepresentation, which results if the assessment is too narrow and does not capture the essential aspects of the construct at hand. Drawing on Messick's explanations on threats to the construct validity, Schellings (2011) recommends that questionnaires be more tailored to the target task in order to reduce the construct-irrelevant information. He goes on to advise that meticulously constructed and examined task-

specific questionnaires may be an adequate alternative method for the think-aloud method and may result in higher correlations in multi-method studies.

To achieve the same purpose, a series of questionnaires referred to in this study as Metacognitive Strategy Questionnaire (MSQ), were devised and administered. These questionnaires provided corroborating data for the think-alouds and enabled objective analysis of pre-to-post-test perceptions of metacognitive strategies specific to the domain of academic writing. In particular, they helped investigate to what degree the participants considered themselves to be planners, considerers of the intended audience, monitors, and evaluators of their essay writing. These questionnaires were specifically used as a measure to tackle the incompleteness of think-aloud protocols, given that, as Hayes and Nash (1996) have rightly claimed, people "cannot articulate everything that crosses their minds" (p. 45). The questionnaires were accordingly intended to help gain as much data as possible to answer the fourth research question:

> 4. What metacognitive strategies do L2 learners perceive in their actual writing performance?

The Metacognitive Strategy Questionnaire (MSQ) was constructed on the basis of Flavell et al.'s (2002) taxonomy of the superordinate metacognitive strategies. These strategies adhered to planning, monitoring, and evaluating. According to Schraw and Moshman (1995, pp. 354-355), planning refers to "the selection of appropriate strategies and the allocation of resources that affect performance." Monitoring involves "one's online awareness of comprehension and task performance." Evaluation refers to "appraising the products and regulatory processes of one's learning. Similar to the coding of think-aloud protocols, another strategy named as considering the audience was added. The questionnaire was a task-specific one, aimed at measuring the same metacognitive strategies that were instructed throughout the metacognitive treatment sessions and were similarly coded in the think-aloud protocols. These strategies formed the basis of the formulation of questionnaire items. All items were phrased and formulated in a way that they explicitly referred to the task at hand. In the pre-test, immediate post-test and delayed post-test, the questionnaires were identical and consisted of 20 items. The questionnaire comprised a number of items: planning (5 items), considering the audience (3 items), monitoring (8 items), and evaluating (4 items). By checking a box on a six-point scale, the participants indicated the extent to which each statement in the preceding writing task

performance pertained to them. Every questionnaire received a score for each of the four dimensions as well as a total score for the whole questionnaire.

6.2.1. Results

In order to compare participants' responses to the items in the questionnaires, they were subjected to statistical analysis. Analyses at item level provided an opportunity to scrutinise whether participants had reported their writing activities differently over time. In order to obtain a summated scale, the final score for the participants on each scale which was the sum of their ratings for the items in each four category was calculated.

Both Boone Jr. and Boone (2012), and Clason and Dormody (1994) differentiate between Likert-type items and Likert scales. They identified Likert-type items as single items that use some aspect of the original Likert response alternatives. On the other hand, a Likert scale is composed of a series of Likert-type items which are combined into a single composite score/variable. Because of these conditions, they explain, Likert-type items fall into the ordinal measurement scale, whereas the Likert scale data are analysed at the interval measurement scale. Based on this description, I decided that the questionnaire data analysis in my study is based on the composite score from the series of items that represent different categories. In other words, while twenty items were used in this questionnaire instrument, the individual items were not intended to be analysed separately and the main attempt was to combine the responses from the items into four composite scales. The items combined into four categories were therefore used to provide a quantitative measure of the metacognitive strategies. Analysing the variables required analysing frequencies, i.e. the number of occurrences that fall into each combination of categories. Furthermore, as reported in the methodology chapter, the Cronbach's alpha for the four different dimensions were as follows: $\alpha = .6$ for planning items, $\alpha = .61$ for considering the audience items, $\alpha = .69$ for monitoring items, and $\alpha = .62$ for evaluating items. Having obtained such internal consistency reliabilities, I perceived it justifiable to aggregate the items per dimension and calculate four mean scores per participants' questionnaires. The same procedure was followed in the studies of Kieft, Rijlaarsdam, and van den Bergh (2006, 2008), and Tillema et al. (2011). In this way, the means of the constituent items focusing on each content area were summed up in multi-item scales. To present the analysis of the data, the descriptive statistics for these interval scale items

including the mean for central tendency and standard deviations for variability is first reported. Table 6.10 below features the descriptive information for each four scale and for each four group.

Table 6.10 Descriptive statistics: MSQ

Group	Questionnaire scale		Pre-test	Immediate Post-test	Delayed Post-test
		N items	M (SD)	M (SD)	M (SD)
+M+P	Planning	5	4.38 (.76)	5.00 (.45)	5.18 (.28)
	Considering the audience	3	4.09 (.75)	5.02 (.51)	4.90 (.27)
	Monitoring	8	4.44 (.70)	4.87 (.59)	4.65 (.22)
	Evaluating	4	4.50 (.51)	4.55 (.41)	4.62 (.54)
+M-P	Planning	5	4.32 (.42)	4.83 (.80)	4.70 (.38)
	Considering the audience	3	3.97 (.98)	5.05 (.63)	4.97 (.34)
	Monitoring	8	4.45 (.79)	5.29 (.53)	4.75 (.33)
	Evaluating	4	4.52 (.26)	4.67 (.34)	5.00 (.27)
-M+P	Planning	5	4.42 (.40)	5.01 (.25)	4.91 (.35)
	Considering the audience	3	4.42 (.56)	4.35 (.44)	4.64 (.63)
	Monitoring	8	4.60 (.32)	4.61 (.37)	4.47 (.25)
	Evaluating	4	4.66 (.42)	4.48 (.42)	4.58 (.36)
-M-P	Planning	5	4.58 (.64)	4.55 (.44)	4.55 (.46)
	Considering the audience	3	4.40 (.73)	4.57 (.44)	4.42 (.35)
	Monitoring	8	4.44 (.37)	4.56 (.34)	4.47 (.23)
	Evaluating	4	4.37 (.41)	4.60 (.42)	4.60 (.49)

Although calculating the mean of these variables can provide useful information, it does not have the power to differentiate between groups. Therefore, after a composite score (mean) from a series of Likert-type items in each category was created, additional data analysis procedures appropriate for interval scale items including ANOVA and repeated measures analysis were then carried out. In what follows, the analysis of each of the four categories is presented separately.

6.2.1.1. Planning

One of the categories affected by metacognitive instruction was planning. As presented in Figure 6.4, the mean scores for the planning category show that the four groups reported somewhat similarly in the pre-test (the mean ranging from 4.32 to 4.58). In the immediate

post-test, however, the average mean scores of the two planning groups, that is to say Group +M+P and Group -M+P, were higher (M = 5.00). In the delayed post-test, the mean scores for all groups were almost the same as in the immediate post-test, except for Group +M+P which reached the highest point of 5.18.

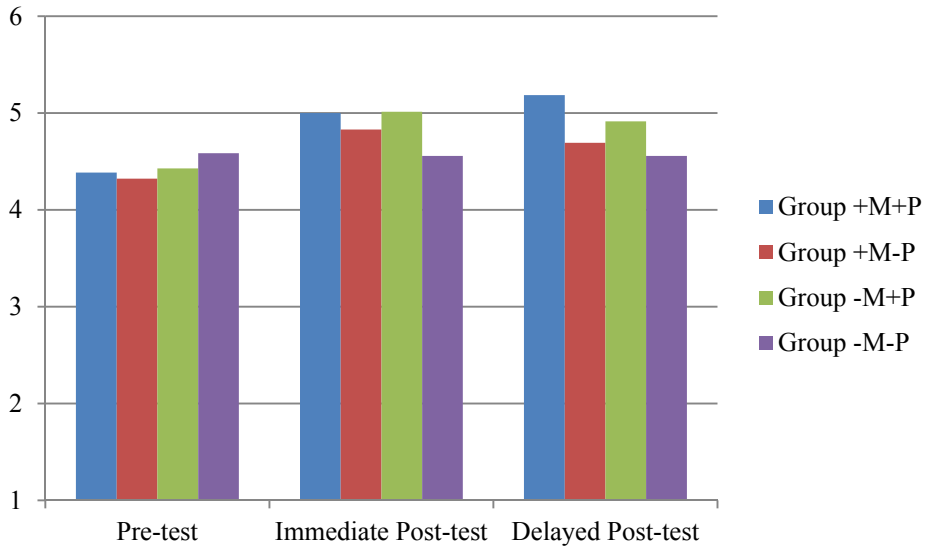


Figure 6.4 Planning in MSQ

Note. M: Metacognitive instruction, P: Planning

In order to specify whether these differences across groups were statistically significant, a one-way ANOVA was calculated. As presented in Table 6.11, only in the delayed post-test were there significant differences among groups (p = .00, large effect size of d = .55).

Table 6.11 ANOVA: Planning

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	.51	3	.17	.50	.67
	Within Groups	17.22	51	.33		
	Total	17.74	54			
Immediate Post-test	Between Groups	1.90	3	.63	2.32	0.08
	Within Groups	13.88	51	.32		
	Total	15.77	54			
Delayed Post-test	Between Groups	3.15	3	1.05	7.36	.00*
	Within Groups	7.27	51	.14		
	Total	10.43	54			

*. The significance is at the .05 level

In order to identify where the differences lay in the delayed post-test, post-hoc Tukey test analysis was calculated. As shown in Table 6.12, in the delayed post-test, Group +M+P showed higher level of agreement with the items in the planning category than Groups +M-P (p = .00, d = .40) and -M-P (p = .00, d = .64). As well as being statistically significant, the effect sizes are medium for the former and large for the latter, and so represent a substantive finding. Such results are not surprising because neither of the non-planning groups, whether they had received the metacognitive instruction or not, were given the opportunity for planning prior to their writing composition.

Table 6.12 Tukey: Planning (Delayed post-test)

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
	+M-P	.49*	.14	.00
+M+P	-M+P	.27	.14	.24
	-M-P	.62*	.14	.00
	+M+P	-.49*	.14	.00
+M-P	-M+P	-.22	.14	.43
	-M-P	.13	.14	.79
	+M+P	-.27	.14	.24
-M+P	+M-P	.22	.14	.43
	-M-P	.35	.14	.07
	+M+P	-.62*	.14	.00
-M-P	+M-P	-.13	.14	.79
	-M+P	-.35	.14	.07

Note. M: Metacognitive instruction, P: Planning
* The mean difference is significant at the .05 level.

Changes in participants' responses over time were calculated by repeated-measures ANOVA. A Bonferroni correction was selected as a more robust technique. In order to indicate whether the assumption of sphericity was met, the Mauchly's test was also calculated. The results of within-participant variance showed that there existed significant differences across the pre-test to the (immediate/delayed) post-tests, F(1.72, 93.04) = 15, p = .00. The post hoc analysis showing the pairwise comparison (Table 6.13) indicated that Group +M+P changed significantly in their responses both from the pre-test to the immediate post-test (p = .01) and from the pre-test to the delayed post-test (p = .00). The same pattern was observed for the non-metacognitive group (with planning), i.e. Group -M+P (p = .01 for both comparisons). Within the no planning groups, Group -M-P did not

show any change over time, but Group +M-P advanced substantially in their perceptions towards planning in the delayed post-test compared to the pre-test (p = .02). This is probably because although they were not allowed to pre-plan prior to their task performance, they were instructed on the significant role of planning, and thus had acquired the awareness.

Table 6.13 Pairwise comparisons: Planning

	(I) Lexical Variety	(J) Lexical Variety	Mean Difference (I-J)	Std. Error	Sig.
+M+P	Pre-test	Immediate Post-test	-.61*	.16	.01
		Delayed Post-test	-.80*	.16	.00
	Immediate Post-test	Pre-test	.61*	.16	.01
		Delayed Post-test	-.18	.08	.11
	Delayed Post-test	Pre-test	.80*	.16	.00
		Immediate Post-test	.18	.08	.11
+M-P	Pre-test	Immediate Post-test	-.50	.20	.08
		Delayed Post-test	-.37*	.11	.02
	Immediate Post-test	Pre-test	.50	.20	.08
		Delayed Post-test	.13	.13	1.00
	Delayed Post-test	Pre-test	.37*	.11	.02
		Immediate Post-test	-.13	.13	1.00
-M+P	Pre-test	Immediate Post-test	-.58*	.10	.00
		Delayed Post-test	-.48*	.11	.00
	Immediate Post-test	Pre-test	.58*	.10	.00
		Delayed Post-test	.10	.11	1.00
	Delayed Post-test	Pre-test	.48*	.11	.00
		Immediate Post-test	-.10	.11	1.00
-M-P	Pre-test	Immediate Post-test	.03	.20	1.00
		Delayed Post-test	.03	.26	1.00
	Immediate Post-test	Pre-test	-.03	.20	1.00
		Delayed Post-test	.00	.16	1.00
	Delayed Post-test	Pre-test	-.03	.26	1.00
		Immediate Post-test	.00	.16	1.00

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

6.2.1.2. Considering the audience

The second category in the questionnaire pertains to considering the audience. In the pre-test, the two metacognitive groups rated the items relating to this strategy almost equally

(approximately 4), whereas their ratings in the immediate and delayed post-tests surpassed 5. One can argue that this increase in rating was a consequence of the metacognitive instruction. The non-metacognitive groups showed a different pattern (see Figure 6.5). Their possible considerations of the audience remained unchanged throughout the experiment. This however was expected, since these two groups did not receive any instruction on how to perceive their audience in their composition.

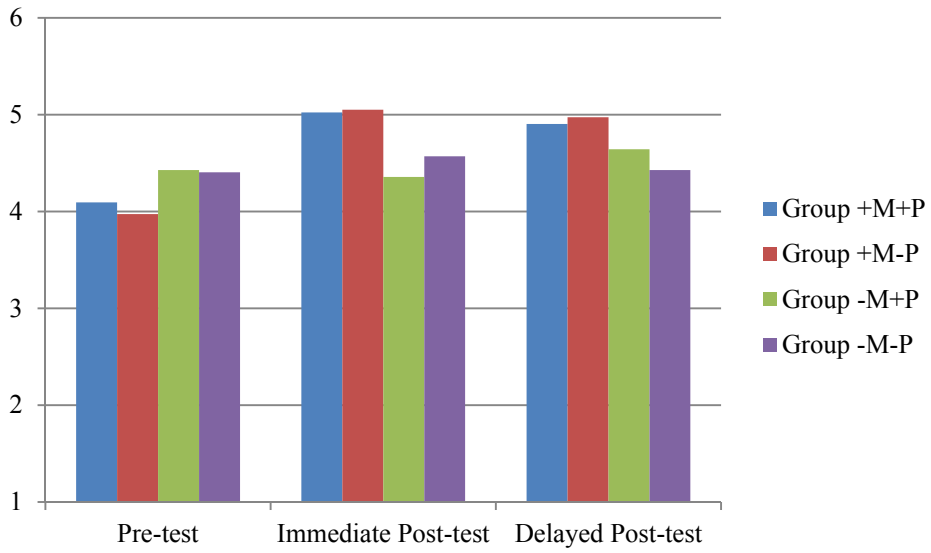


Figure 6.5 Considering the audience in MSQ

Note. M: Metacognitive instruction, P: Planning

As shown in Table 6.14, the results of the one-way ANOVA analysis demonstrated that there were significant across-group differences in participants' responses to considering the audience items in both immediate post-test ($p = .00$, large effect size of $d = .51$) and delayed post-test ($p = .00$, medium effect size of $d = .46$).

Table 6.14 ANOVA: Considering the audience

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	2.08	3	.69	1.17	.32
	Within Groups	30.15	51	.59		
	Total	32.23	54			
Immediate Post-test	Between Groups	4.81	3	1.60	6.12	.00*
	Within Groups	13.38	51	.26		
	Total	18.19	54			
Delayed Post-test	Between Groups	2.58	3	.86	4.73	.00*
	Within Groups	9.28	51	.18		
	Total	11.86	54			

*. The significance is at the .05 level

The post-hoc Tukey test analysis (Table 6.15) indicated that in the immediate post-test, Group +M+P (p = .00, d = .58) and Group +M-P (p = .00, d = .55) agreed to have considered the audience in a greater extent than Group -M+P. In the delayed post-test, again both metacognitive groups rated the questionnaire items relating to considering the audience category higher than Group -M-P (p = .02, d = .61 for the +M+P/-M-P comparison, and p = .01, d = .62 for the +M-P/-M-P comparison).

Table 6.15 Tukey: Considering the audience (Immediate and delayed post-tests)

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
Immediate post-test	+M+P	+M-P	-.02	.19	.99
		-M+P	.66*	.19	.00
		-M-P	.45	.19	.10
	+M-P	+M+P	.02	.19	.99
		-M+P	.69*	.19	.00
		-M-P	.47	.19	.08
	-M+P	+M+P	-.66*	.19	.00
		+M-P	-.69*	.19	.00
		-M-P	-.21	.19	.68
	-M-P	+M+P	-.45	.19	.10
		+M-P	-.47	.19	.08
		-M+P	.21	.19	.68
Delayed post-test	+M+P	+M-P	-.06	.16	.97
		-M+P	.26	.16	.37
		-M-P	.47*	.16	.02

		+M+P	.06	.16	.97
	+M-P	-M+P	.33	.16	.19
		-M-P	.54*	.16	.01
		+M+P	-.26	.16	.37
	-M+P	+M-P	-.33	.16	.19
		-M-P	.21	.16	.54
		+M+P	-.47*	.16	.02
	-M-P	+M-P	-.54*	.16	.01
		-M+P	-.21	.16	.54

Note. M: Metacognitive instruction, P: Planning

* The mean difference is significant at the .05 level.

In order to ascertain the within-group changes in participants' perception to the use of strategies over time, repeated-measures ANOVA was calculated. The results confirm that there existed significant differences from the pre-test to the immediate post-test to the delayed post-test, $F(1.71, 92.66) = 14.43$, $p = .00$. Table 6.16 shows the post hoc tests for the repeated-measures analysis of the metacognitive and non-metacognitive groups. Both metacognitive groups increased in their rating of considering the audience strategy. This was significant not only in the immediate post-test compared to the pre-test ($p = .00$ for Group +M+P; and $p = .00$ for Group +M+P), but also in the delayed post-test compared to the pre-test ($p = .00$ for Group +M+P; and $p = .00$ for Group +M+P). There were, however, no significant differences across the pre-test to the two post-tests in either of the non-metacognitive groups.

Table 6.16 Pairwise comparisons: Considering the Audience

	(I) Lexical Variety	(J) Lexical Variety	Mean Difference (I-J)	Std. Error	Sig.
	Pre-test	Immediate Post-test	-.92*	.19	.00
		Delayed Post-test	-.81*	.21	.00
+M+P	Immediate Post-test	Pre-test	.92*	.19	.00
		Delayed Post-test	.11	.15	1.00
	Delayed Post-test	Pre-test	.81*	.21	.00
		Immediate Post-test	-.11	.15	1.00
	Pre-test	Immediate Post-test	-1.07*	.16	.00
		Delayed Post-test	-1.00*	.22	.00
	Immediate Post-test	Pre-test	1.07*	.16	.00
+M-P		Delayed Post-test	.07	.16	1.00

	Delayed Post-test	Pre-test	1.00*	.22	.00
		Immediate Post-test	-.07	.16	1.00
-M+P	Pre-test	Immediate Post-test	.07	.19	1.00
		Delayed Post-test	-.21	.19	.86
	Immediate Post-test	Pre-test	-.07	.19	1.00
		Delayed Post-test	-.28	.21	.58
	Delayed Post-test	Pre-test	.21	.19	.86
		Immediate Post-test	.28	.21	.58
-M-P	Pre-test	Immediate Post-test	-.17	.27	1.00
		Delayed Post-test	-.02	.22	1.00
	Immediate Post-test	Pre-test	.17	.27	1.00
		Delayed Post-test	.14	.13	.86
	Delayed Post-test	Pre-test	.02	.22	1.00
		Immediate Post-test	-.14	.13	.86

Note. M: Metacognitive instruction, P: Planning
*. The significance is at the .05 level

6.2.1.3. Monitoring

The third category solicits information on participants' perceptions of having monitored their essay writing. In the pre-test, all four groups rated the items almost identically. In the immediate post-test, however, the average rating of Group +M-P was considerably higher than that of the other groups (M = 5.28). The next highest rating belonged to Group +M+P with a mean of 4.86. In the delayed post-test, however, the trend was not maintained to that extent for Group +M-P, but they still appeared to have the highest rating of all.

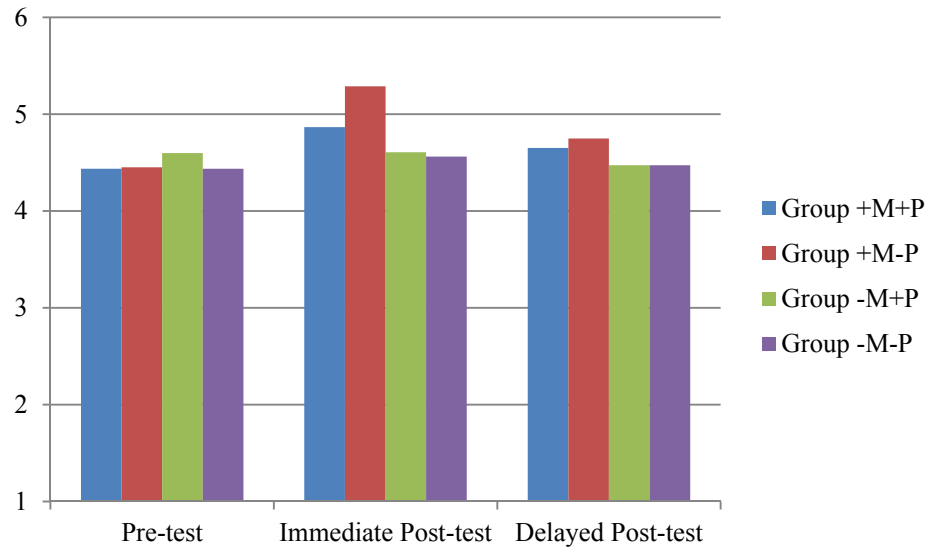


Figure 6.6 Monitoring in MSQ

Note. M: Metacognitive instruction, P: Planning

According to Table 6.17, there were no significant differences among groups in the pre-test. There existed, however, a significant difference among groups in the immediate and delayed post-tests (p = .00, d = .53, and p = .02, d = .42 respectively).

Table 6.17 ANOVA: Monitoring

		Sum Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	.25	3	.08	.25	.86
	Within Groups	17.24	51	.33		
	Total	17.50	54			
Immediate Post-test	Between Groups	4.43	3	1.47	6.63	.00*
	Within Groups	11.36	51	.22		
	Total	15.79	54			
Delayed Post-test	Between Groups	.76	3	.25	3.65	.02*
	Within Groups	3.55	51	.07		
	Total	4.31	54			

*. The significance is at the .05 level

Tukey LSD post-hoc test (Table 6.18) revealed that after the composition stage in the immediate post-test, participants in Group +M-P agreed they had monitored their compositions in a more considerable degree than Group -M+P (p = .00, d = .61) and Group -M-P (p = .00, d = .64). In the delayed post-test, a similar picture emerged for Group +M-P, compared to Group -M+P (p = .04, d = .43) and Group -M-P (p = .04, d = .44). As reported, the effect sizes for the immediate post-test were large, and for the delayed post-test were medium.

Table 6.18 Tukey: Monitoring (Immediate post-test and delayed post-test)

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.
Immediate post-test	+M+P	+M-P	-.42	.18	.10
		-M+P	.26	.17	.47
		-M-P	.30	.17	.33
	+M-P	+M+P	.42	.18	.10
		-M+P	.68*	.18	.00
		-M-P	.72*	.18	.00
	-M+P	+M+P	-.26	.17	.47
		+M-P	-.68*	.18	.00
		-M-P	.04	.17	.99
	-M-P	+M+P	-.30	.17	.33
		+M-P	-.72*	.18	.00
		-M+P	-.04	.17	.99
Delayed post-test	+M+P	+M-P	.09	.10	.77
		-M+P	.17	.10	.29
		-M-P	.17	.10	.29
	+M-P	+M+P	.09	.10	.77
		-M+P	.27*	.10	.04
		-M-P	.27*	.10	.04
	-M+P	+M+P	-.17	.10	.29
		+M-P	-.27*	.10	.04
		-M-P	.00	.10	1.00
	-M-P	+M+P	-.17	.10	.29
		+M-P	-.27*	.10	.04
		-M+P	.00	.10	1.00

Note. M: Metacognitive instruction, P: Planning

* The mean difference is significant at the .05 level.

The results of the repeated-measure ANOVA showed that there were significant differences within groups across time, F(1.80, 97.17) = 9.68, p = .00. As shown in Table 6.19, only the pairwise comparisons for the metacognitive group (without planning) showed both a significant increase in rating from the pre-test to the immediate post-test (p = .00), and a substantial change from the immediate post-test to the delayed post-test (p = .00). Other comparisons in other three groups, however, did not reveal any changes in participants' perceptions over time.

Table 6.19 Pairwise comparisons: Monitoring

	(I) Lexical Variety	(J) Lexical Variety	Mean Difference (I-J)	Std. Error	Sig.
+M+P	Pre-test	Immediate Post-test	-.43	.23	.25
		Delayed Post-test	-.21	.20	.94
	Immediate Post-test	Pre-test	.43	.23	.25
		Delayed Post-test	.21	.14	.50
	Delayed Post-test	Pre-test	.21	.20	.94
		Immediate Post-test	-.21	.14	.50
+M-P	Pre-test	Immediate Post-test	-.83*	.11	.00
		Delayed Post-test	-.29	.18	.40
	Immediate Post-test	Pre-test	.83*	.11	.00
		Delayed Post-test	.53*	.11	.00
	Delayed Post-test	Pre-test	.29	.18	.40
		Immediate Post-test	-.53*	.11	.00
-M+P	Pre-test	Immediate Post-test	-.01	.09	1.00
		Delayed Post-test	.12	.13	1.00
	Immediate Post-test	Pre-test	.01	.09	1.00
		Delayed Post-test	.13	.14	1.00
	Delayed Post-test	Pre-test	-.12	.13	1.00
		Immediate Post-test	-.13	.14	1.00
-M-P	Pre-test	Immediate Post-test	-.12	.15	1.00
		Delayed Post-test	.03	.12	1.00
	Immediate Post-test	Pre-test	.12	.15	1.00
		Delayed Post-test	.09	.09	1.00
	Delayed Post-test	Pre-test	.03	.12	1.00
		Immediate Post-test	-.09	.09	1.00

Note. M: Metacognitive instruction, P: Planning

*. The significance is at the .05 level

6.2.1.4. Evaluating

Evaluating the writing essay was also a dominant strategy under the instructional lesson planning. In MSQ, it was targeted by four items in the last category. As shown in Figure 6.7, all four groups appeared to have rated the items almost identically in the pre-test and immediate post-test. In the delayed post-test, however, the average rating of Group +M-P was noticeably higher than that of other groups (M = 5.05).

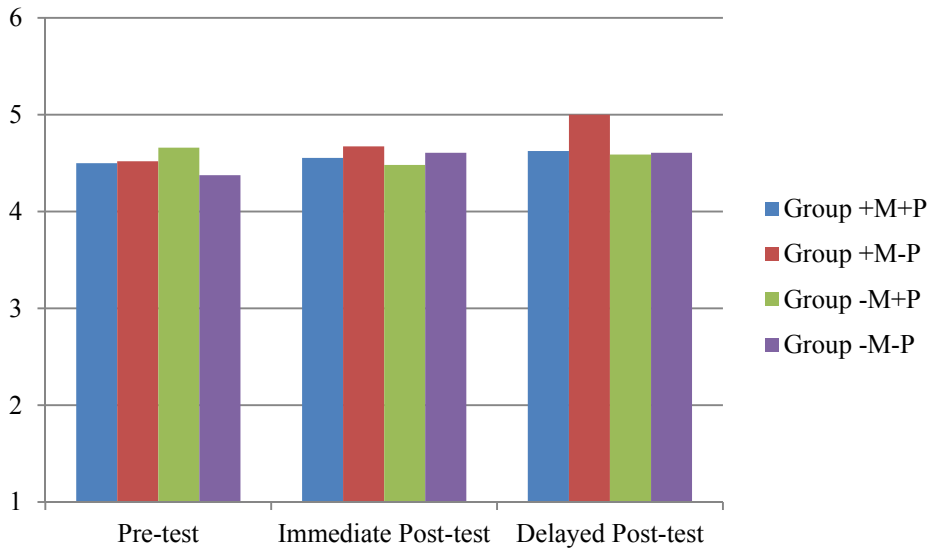


Figure 6.7 Evaluating in MSQ

Note. M: Metacognitive instruction, P: Planning

The ANOVA analysis showed that there were no significant differences among groups in terms of participants' perceptions of having evaluated their essays in the pre-test or the post-tests. In the delayed post-test, the significance showed a tendency to significance ($p = .054$, $d = .37$), but as reported the effect size was medium. To ensure this tendency resulted in any differences among groups in the delayed post-test, the Tukey test was run but no significant differences were detected.

The results of within-group variance showed that there existed significant differences across the pre-test to the immediate and delayed post-tests, $F(1.98, 107.22) = 2.79$, $p = .00$. Table 6.20 shows the post-hoc pairwise comparisons of groups over time. It was only Group +M-P which changed in their perceptions of having evaluated their compositions in

the delayed post-test compared to the pre-test (p = .02), and in the delayed post-test compared to the immediate post-test (p = .00).

Table 6.20 Pairwise comparisons: Evaluating

	(I) Lexical Variety	(J) Lexical Variety	Mean Difference (I-J)	Std. Error	Sig.
+M+P	Pre-test	Immediate Post-test	-.05	.19	1.00
		Delayed Post-test	-.12	.20	1.00
	Immediate Post-test	Pre-test	.05	.19	1.00
		Delayed Post-test	.07	.19	1.00
	Delayed Post-test	Pre-test	.12	.20	1.00
		Immediate Post-test	.07	.19	1.00
+M-P	Pre-test	Immediate Post-test	-.15	.11	.62
		Delayed Post-test	-.48*	.10	.00
	Immediate Post-test	Pre-test	.15	.11	.62
		Delayed Post-test	-.32*	.07	.00
	Delayed Post-test	Pre-test	.48*	.10	.00
		Immediate Post-test	.32*	.07	.00
-M+P	Pre-test	Immediate Post-test	.18	.18	1.00
		Delayed Post-test	.07	.15	1.00
	Immediate Post-test	Pre-test	-.18	.18	1.00
		Delayed Post-test	-.10	.14	1.00
	Delayed Post-test	Pre-test	-.07	.15	1.00
		Immediate Post-test	.10	.14	1.00
-M-P	Pre-test	Immediate Post-test	-.23	.15	.48
		Delayed Post-test	-.23	.18	.66
	Immediate Post-test	Pre-test	.23	.15	.48
		Delayed Post-test	.00	.14	1.00
	Delayed Post-test	Pre-test	.20	.18	.66
		Immediate Post-test	.00	.14	1.00

Note. M: Metacognitive instruction, P: Planning

*. The significance is at the .05 level

Summary

On the whole, the results obtained from the questionnaire data were consistent with the findings of the quantitative textual data vis-à-vis the instructional differences among groups. A comparison between the questionnaire results and those of the text analysis indicated an advantage for the participants engaged in pre-task planning over the no planning condition in responding to the planning items. Metacognitive instruction was

found to have a significant impact on participants' responses to items in considering the audience category, whereas it was found to have no effect for other non-metacognitively instructed groups. Examining the comparison between groups with regard to the monitoring and evaluating variables, it was noticed that the differences reached statistical significance for the metacognitive group (without planning). This group showed to have devoted more attention to the execution and monitoring aspects of the composing process.

Conclusion

Overall, the actual writing texts produced by the participants constituted the main data for this project. However, the transcripts of the think-aloud protocols and the participants' responses to MSQ in the pre-test, immediate and delayed post-tests were additionally analysed in order to help interpret the findings of the statistical analysis of the main source of quantitative data. Some volunteer participants in Study 2 were required to access their memory for the writing activities they used during the task execution and to give retrospective accounts on how they performed the task over multiple occurrences of writing. According to Schellings (2011), when the respondents to a questionnaire experience difficulty interpreting an item, they tend to choose the neutral answer in odd-numbered Likert-scale questionnaires, or an answer alternative in the middle of the scale in even-numbered questionnaires. A similar pattern was observed in my study. In the pre-test, the mean scores of the scales in all groups were approximately between 4 and 4.50. The participants may have been rating their general approach to writing, rather than the specific task they performed. In addition, even if the participants may not have understood the true intention of an item or may have been uncertain in recollecting whether they truly performed the task as questioned in the items, they may have had the tendency to answer the questions in a manner that would be viewed more favourably by different persons the respondents compared themselves to, including the test administrator, the teacher, or the best/worst student in class. This is what Schellings (2011) and Schellings et al. (2012) call answering the questions from a social desirability perspective. Also, as claimed by Schellings (2011), when reporting on their conceptions of the activities they performed, respondents may tend to choose a stable response pattern. In a similar vein, the results of the Likert-scale Questionnaires demonstrated that participants, on average, chose one or two reference points when reporting their conceptions about their performance. For instance, mainly Group +M+P and to some extent Group -M+P offered strongly agreed

responses under the planning category, while other groups mostly offered agreed responses in this respect.

CHAPTER 7. CONCLUSION

This chapter presents a summary of key findings followed by a consideration of the theoretical and pedagogical implications of this study. Its limitations are assessed and possible directions for further research are suggested.

7.1. Review of the study

This study examined different kinds of writing instruction delivered to ESL learners in Study 1 and EFL learners in Study 2. Data analysis was carried out from cognitive and psycholinguistic perspectives. Two theoretical strands underpinned the study. Theories of writing research from the broad perspectives of *texts*, *writers*, and *readers* provided insights into the research design and methods of the study, and Skehan and Foster's (2001) Limited Attentional Capacity Model was used to discuss the critical issues in relation to writing practices.

Two studies were conducted: in a university in Australia (n = 35), and in a language school in Iran (n = 70). Two types of intervention (i.e. metacognitive strategy instruction and writing instruction) were implemented with participants from these two target populations. Quantitative and qualitative methods were utilised to collect and analyse the data. Four sessions of instructional intervention (8 hours) were conducted, along with separate sessions in which students participated in a pre-test, immediate post-test (in Study 1 and 2), and delayed post-test (in Study 2). Drawing upon Anderson's (2007) claim that "not all writers approach a writing task in exactly the same way" (p. 20), it was considered beneficial to the findings of the project to look beyond the written texts that participants produced. In Study 2, a series of Metacognitive Strategy Questionnaire were administered to the four experimental groups in the pre-test and post-tests. Also, the relative frequencies and differing patterns of metacognitive strategy-related verbalisations in six volunteer participants' concurrent data were established. This way, the multiple approaches that different participants adopted to think through the solutions to a writing task became more visible.

Study 1 functioned as a linked study which was undertaken initially to enable the researcher to indicate gaps, to familiarise with the research procedures to follow in Study 2, and to allow adaptations in data collection, data organisation and data analysis. On the basis of the outcomes of Study 1, the researcher ensured the research design was feasible to carry out, the research questions were plausible to answer, the provided scope of the research was manageable, the materials, instruments, writing topics, questionnaires and writing models and samples were dependable, and the theoretical framework, adopted and developed out of combining elements from several theoretical frameworks, was appropriate.

The first research question explored the effectiveness of metacognitive instruction and pre-task planning in terms of general writing proficiency. In Study 1, the combination of the two variables resulted in higher-quality written products over time. In Study 2, although all groups significantly progressed over time, it was the metacognitive groups that showed better performance in the delayed post-test in comparison to the control group and the non-metacognitive (without planning) group. Interestingly, an increase in sample size strengthened the significance of metacognitive instruction.

In relation to the second research question as regards the accuracy of written products, except for the metacognitive (without) planning group, all the other three groups improved in the post-test in producing texts which contained error-free clauses. A different picture, however, emerged in Study 2. As in the first study, there were no differences between groups, but with the exception of group +M+P, none of the groups improved in producing texts containing accurate clauses on the post-tests. Concerning the accuracy of verb forms, there was no evidence of any group differences when participants in both studies performed the tasks over time. Therefore, the results were similar according to the context and neither of the instructions had any impacts on the accuracy of verb forms.

With respect to the third research question in terms of the complexity of written products, none of the complexity measures in Study 1 showed differences between groups, except for syntactic variety in which the metacognitive (with planning) group outperformed the non-metacognitive (without planning) group. In Study 2, the metacognitive groups produced texts with higher lexical variety in the immediate post-test compared to the control group and non-metacognitive (without planning) group. This more skilled

performance was not, however, observed for syntactic complexity or syntactic variety. This shows that the nature of the essays and the type of instruction that metacognitive groups received predisposed them to make decisions about how to allocate their attentional resources by prioritizing the lexical aspect of language over others.

Following the textual analysis of written products, the results of the think-aloud protocol analysis further suggested that although participants in the non-metacognitive groups did have some metacognitive awareness to plan, consider the audience, monitor, and evaluate their writing performance, it was not as sufficient or as successful as those in the metacognitive groups. In addition, the findings from the Metacognitive Strategy Questionnaire revealed that the participants in Group +M+P, and Group +M-P shared a rather similar orientation towards their writing outputs in terms of the subscales of the questionnaire. Compared to the two no-planning groups, the metacognitive (with planning) group showed greater agreement with the planning items of the questionnaire in the delayed post-test. Having been instructed that one of the primary purposes of an argumentative essay is to convince the reader of the validity of one's point of view, participants in the metacognitive groups both rated the items in the second subscale relating to considering the audience higher than the non-metacognitive groups. Concerning the monitoring subscale of the questionnaire, the non-metacognitive (without planning) group agreed they had monitored their writing more than the metacognitive groups. This was most likely because they had no opportunity to plan prior to the task and were therefore obliged to attend to all aspects of the task while producing the written argument. Finally, because all groups were pressured to finish in forty five minutes, they had difficulty attending to the evaluation of their written products, which is probably why no differences were detected among participants' rating of evaluation items of the questionnaire.

7.2. Research implications

This section discusses the relevance of the findings of the study to the theoretical knowledge in this area and to L2 writing pedagogy.

7.2.1. Theoretical implications

From a theoretical perspective, some interplay and trade-off effects as proposed by Skehan and Foster (1997, 2001) were observed in this study. The finding that focusing on lexical complexity tended to be at the expense of accuracy suggests that the cognitively demanding tasks of academic writing directed learners' attention more to meaning than form. Therefore, the linguistic well-formedness of performance was not attended to as much as the lexical richness of the written products. This resulting perspective is in need of further contemplation if it is to help advance the usefulness of planning in the writing process. For teaching practice this implies that promoting both accuracy and complexity may not be easily achieved. However, it is possible to design tasks which direct learners' attention to specific linguistic features or particular task aspects (e.g. Ellis, 2008; Long, 2000).

Secondly, from an interactionist-cognitive perspective, it can be argued that when writers are given explicit instruction on metacognitive strategies, they tend to develop a greater level of awareness and more explicit knowledge of writing. The theoretical assumption here is that explicit knowledge can then facilitate processes such as understanding and noticing which are considered to be effective in helping develop implicit knowledge (Schmidt, 1994). In this study, only the accuracy of writing was shown to be compromised in argumentative writing tasks, which appeared to impose more attentional demands on writers. Previous studies have examined consciousness-raising tasks, particularly for the acquisition of grammar, within the compass of interactionist-cognitive theory (e.g. Eckerth, 2008; Fotos, 1993, 1994; Fotos & Ellis, 1991). These studies have provided evidence that such tasks help learners to develop explicit knowledge in L2 which may be subsequently responsible for the acquisition of implicit knowledge. Likewise, Shin (2008) recommends that L2 writing instructors organize consciousness-raising activities for various genres by explicitly and systematically explaining the ways language functions in social and educational contexts (p. 128).

Thirdly, the study makes a novel contribution to the research of L2 writing in that it adopted an interactionist position and provided a socio-cognitive perspective on addressing the complex interrelationships between approaches to writing in terms of *texts*, *writers*, and *readers*. This study provided an example of taking a contextualised, process-

oriented approach to the design of an L2 writing intervention. In doing so, it was able to investigate the influences of training on learners' writing development (*text-oriented*), the learners' cognitive performances through their on-task verbal reports (*writer-oriented*), and the learners' responses to a particular communicative setting (*reader-oriented*). While most previous writing research studies generally adhere to one approach, the present study took a step further to investigate synergies and interdependence between these three dimensions.

Finally, in Hartman and Sternberg's (1993) BACEIS model of the interaction between cognitive, metacognitive, and affective components of learning, there is an external supersystem consisting of the environmental context. This supersystem includes both an academic (classroom, curricular, and teacher characteristics), and a non-academic system (cultural, economic, and familial factors). The complexity of this model demonstrates that metacognition functions within a complex system in which cognitive and affective factors interact with context-dependent issues. Similarly, Hayes' (1996) model of the writing process consists of two major components. The first is the task environment, which includes both the social and physical environment, and the second is the individual, which incorporates motivation/affect, cognitive processes, working memory and long-term memory. Consequently, instructors are advised to take into consideration how to design and implement tasks within a wider framework of real classroom contexts to secure a more detailed and comprehensive teaching and learning environment.

7.2.2. Pedagogical implications

A number of pedagogical implications arise from the findings of the present study for teachers' reflection and classroom practice. Firstly, the main purpose of the current instructional intervention was to teach metacognitive strategies pertaining to academic writing, but also to help participants to construct explicit knowledge regarding when, where, how and why to use these strategies. It is important to stress that the metacognitive intervention in this study followed an interactive approach and consisted of a variety of instructional practices such as direct instruction, teacher modelling, reflection on the part of participants, and group discussions, which all allowed them to share their knowledge of strategies. As the researcher in this study, I made constant effort in the metacognitively instructed condition to model my own metacognition for the participants. To be more

specific, in addition to discussing and modelling how to perform argumentative writing tasks, I described the thought processes required when writing the essay (i.e. how to think about, plan and monitor the performance). Based on this interactive approach to teaching L2 writing, the results suggested that once participants were engaged in pre-task planning activities and were trained metacognitively, they yielded texts with a consequential number of error-free clauses and a richer range of vocabulary.

Additionally, it is generally acknowledged that the classroom instruction should not follow a one-size-fits-all approach. Instead, it needs to be modified so that it is effective for most, not ideally all, students. A sociocultural perspective on writing instruction (e.g. Englert, 1992; Englert et al., 2006) emphasises the need to adjust the type of instruction offered to learners in order to suit their stage of development, as there may not be just one way to instruct learners how to write effectively. In addition to the learners' stage of development, the instruction may well depend upon the learners' attitudes, and the particular contexts within which they find themselves, all of which would influence what type of instruction may be efficient. In the present study, the participants in the non-metacognitive groups also progressed over time both in their general writing proficiency and in some micro-skills of their production. This leads us to the conclusion that genre-based instruction in these groups can also be considered as an effective approach to promote L2 learner's writing ability. When these learners became more familiar with common genres, they seemed to have developed shortcuts to effective processing and production of texts. Teachers therefore need to consider various options, formulate an explicit policy for instructing writing, and subject their policy to evaluation by considering other types of methods (for example through action research). From an instructional practices perspective, what is possible is to adopt an approach to writing instruction which combines explicit instruction in metacognitive strategies, particularly pre-task planning techniques, with the typical features of each genre of writing. Effective writing instruction needs to include teaching L2 learners the metacognitive skills of knowing when and where to use different writing strategies of how to plan, monitor, and control their writing, and how to transfer the skills required in the classroom to other contexts.

As the next pedagogical implication, the findings indicated that scaffolded extended practice facilitates the use of each strategy required for the argumentative mode of discourse in writing. When participants of this study were given regular opportunities for

extended writing practice, and opportunities to reflect at the start of each session on their successes and failures of the previous session writing practice, they gained more confidence in constructing metacognitive knowledge and regulatory skills in a more skilled manner. Learners are not expected to independently develop competence in these strategies. For the same reason, I endorse Goodrich Andrade and Perkins (1998) who argue that students possess some inert knowledge and skills but that they do not think to use them unless they are explicitly instructed to. In a similar vein, Hartman (2001) asserts that these strategies need to be "explicitly and continually addressed, practiced, polished, and internalised" (p. 40). Therefore improvement will come about as a result of both explicit instruction and extended practice. Studies examining ways to promote metacognitive awareness have similarly shown that extended practice and reflection play key roles in the construction of metacognitive knowledge and regulatory skills (e.g. Gourgey, 2001; Kuhn, Schauble, & Garcia-Mila, 1992; Schraw, 2001). Gourgey (2001), for instance, argues that learners can be explicitly instructed to improve metacognitive proficiency through repeated guided practice.

As another implication, the actual concern is how writing instruction can be re-structured to better facilitate the acquisition of higher-order cognitive skills in educational settings. As an insider in the EFL context of Iran and familiar with the audio-lingual and grammar-translation system of English education, my observation was that teaching writing skills to students at secondary and tertiary schools can prove challenging. The mainstream education system is not balanced with regard to its instructional emphasis on language skills. Therefore, when compounded by the disinterest of policy makers in the enhancement of writing skill and the dominance of an exam-oriented education system where reading takes prominence over other skills, the entire exercise of teaching writing becomes more demanding for teachers and educators. In addition, the neglect of teaching writing often serves to reinforce the commonly held conviction that this skill is strictly related to tertiary students who study English majors only. By the time students are at high school or even university in Iran, both speaking and writing skills are considered only peripheral. Writing becomes an increasingly critical tool for fulfilling personal, academic or professional potentials only for those students who take private language classes or who wish to pursue their studies outside Iran and thus are obliged to take an international proficiency exam like IELTS or TOEFL. Such students are aware that if they have not learned how to express themselves in writing, they will not be able to communicate

properly with their professors when writing proposals, reports, research papers, dissertations or theses. The participants in this study all intended to take the academic IELTS exam and had a fixed period of time to prepare themselves. The needs of the participants were therefore different because comparatively more communicative skills were necessary for them to obtain the required score in IELTS exams. They were attending the private sector, because the secondary body of the educational system of the country, as Hayati and Mashhadi (2010) have put it, shoulders the responsibility of helping the public sector to meet the country's rising demand. The immediate implication of this study concerns the need for teachers to shift their conceptions of professionalism from merely being shaped by the demands of nationwide exams to innovatively planning and devising tasks that go beyond lower-order cognitive skills. These tasks can be scaffolded by a motivating and interactive environment which encourages active cooperation of students in classrooms at all levels. In this study, sufficient supports were provided to promote learning when different concepts and skills were first introduced. These supports included different resources, templates and guides. Tasks were modelled, topics were analysed prior to the composition, and concepts were consolidated by referring to previous session writing practices. These supports were gradually removed as participants were expected to have developed autonomous strategies in the post-test writing tasks. As also supported by research studies (e.g. Ryan & Patrick, 2001; Skinner, Kindermann, & Furrer, 2009), a positive social environment in the class, including teacher support, promoting interaction, mutual respect and performance goals, can correlate with the students' expectations of success and can result in academic optimism. Perhaps, for a start, it would be feasible for teachers and educators to progressively step away from the prescribed structure-based textbooks and examine the metacognitive strategy training which has shown in the literature to contribute to self-regulated learning and the acquisition of critical knowledge and skills pertinent to perform various writing tasks.

Furthermore, metacognition helps learners to use knowledge strategically and to plan, monitor, and evaluate their performances during the execution of their tasks more effectively. Rather than merely focusing learners' attention on paragraph development in different genres of essay writing (as it happened in the non-metacognitive groups in the current study), teachers and educators are recommended to foster a thinking culture in classrooms where learners can develop skills and have opportunities to consciously reflect upon what occurs during the writing process. This process can be examined through

introspective methods such as self-report or self-observation. This way, the think-aloud protocol can be utilised not only as a research tool (as it happened in the current study), but also as a pedagogic tool, a notion that is supported by N. J. Anderson (2007). Drawing on his model of metacognition (see Figure 2.1, chapter 2), some of the reflective pedagogical procedures which Anderson encourages writing teachers to implement were utilised in this study. These included:

1. Demonstrating how to think aloud while writing and weaving in examples of *planning*, *using*, *monitoring*, *orchestrating*, and *evaluating* the work.
2. Grouping participants into pairs to take turns thinking aloud and asking questions about the writing strategies.
3. Varying the pair-work activity by asking participants to focus on interpretation of what has been written.
4. Conducting a "think-aloud round robin" activity, in which students write silently for 3-5 minutes, after which they are invited to share the writing strategies they have used.
5. Conducting a "hot seat" activity, in which an individual participant is selected to write a paragraph in front of the class and his/her writing strategies are discussed.

By the same token, improving a curriculum is not actualized only by the teachers of that system, albeit it is still their first-hand classroom experience. My argument here is not to give a lower profile to the significance of teachers who may appear to be falling at the last stage in the process of language curriculum development which includes other stages such as policy determination, syllabus design, and program implementation (Graves, 1996; R. K. Johnson, 1989; Nation & Macalister, 2010). Rather, I contend that central to any realistic and compatible changes to current writing practices is the involvement of an inter-related network of all the factors in the various stages of curriculum planning and development. I would recommend more sophisticated pre-service and in-service training programs, so that teachers, especially the beginning ones, gain the optimal confidence and awareness in engaging students with the examination and analysis of more critical aspects of writing. To this end, the teacher training programs should also include training teachers metacognitively in the first place and not simply imparting "reductive understandings of the syllabus-relevant conceptual and pedagogical tools" (Kramer-Dahl & Chia, 2012, p. 76). Apart from teachers' professional development, changes within policy making, materials development, and syllabus design need to be taken into account. Improving the

textbooks and materials employed in the program are required to be systematically evaluated on the basis of some well-established criteria which may at the same time be linked to the country's macro and micro policies in foreign language education. The call for such measures of change depends, in due course, upon the Ministry of Education's willingness and endorsement to change. Yet, this should not discourage teachers, educators, and researchers to withdraw their impetus for change. Conversely, convincing the policy makers to seek and pursue a need for change has to originate from a good number of related research-based studies in the first place. As it is, we may still require some time before we see a paradigm shift in policy making involving writing instruction in the mainstream classrooms in Iran. In the long run, however, it is to the society's benefits if students are nurtured in an innovative system of education where it metacognitively empowers the students with creativity, analytical abilities and critical thinking skills. On that note, I hope this study contributes a small part to the ongoing impetus for change.

Finally, according to Hayes (1996), "what we write, how we write, and who we write to is shaped by social convention and by our history of social interaction" (p. 5). In the current study, both the ESL participants in a university in Australia and the EFL participants in a language school in Iran were required to write in a language different from their native tongue. In addition, there tend to be some cultural aspects raised when dealing with argumentation, the construction of which may be distinctly different from their native conditioning. This adds to the complexities inherent in applying a second language to the natural expression, highlighting the cultural differences and expectations exposed when having to adopt a second language as if it were their first language. An L1 English writer could give references and evidence to which a native-speaker rater would relate strongly, but the probability that an L2 learner would produce such evidence/references might not seem very high. Hence, the evident lack of high achievement in an L2 writer who is new to a Western style of argument should come as no surprise. Argumentative writing is not a universal genre of writing. Each culture has its own tradition of utilising information and presenting ideas. Modern argumentation theory in general, and argumentative writing in particular have roots in western academia, originally developed from Roman and Greek thinkers like Plato whose approach to argument is drawn from a common well, the knowledge within which is taken for granted as being shared by all academics. Yet, for many L2 learners, the knowledge of this well may lie beyond their academic experience.

A number of studies provide evidence of the variety of patterns of arguments across different cultures (e.g. Arsyad, 1999; Connor, 1987; Gilbert, 2004; Krause & O'Brien, 1999). Thus for L2 learners, a challenge in writing is not only the language development and paragraph organisation, but the cultural influences on academic writing styles. According to Hyland (2003), teachers should help L2 learners to develop the sociocultural schemata which reflect the ways that members of different discourse communities think. He claims this can be achieved "by extending their knowledge of form, process, and content to the discourse communities within which they serve particular purposes" (p. 25). Otherwise, I would speculate that the magnitude of such an approach could be seen as quite overwhelming for L2 learners who find it a challenge to simply come to terms with their second language in the first place. Overall, the point to make here is that L2 writing development is a long process and while instruction may make some differences (as demonstrated in this study), changes in some aspects of writing will only emerge over time – particularly those that, as I pointed out, are tied to a cultural way of expressing opinions.

7.3. Limitations of the study

Despite the potential contributions of this study to writing research, a number of limitations are noted. Firstly, among the limitations, is the number of participants in this study. We know from the central limit theorem that as the sample gets bigger, more specifically in samples of 30 or more, the sampling distribution tends to normality, and thus parametric tests can be conducted for the data analysis. Although the number of participants increased from 35 in Study 1 to 70 in Study 2, increasing the sample size in each group even further could have better confirmed the effects of planning and metacognitive instruction on the quality of writing.

Secondly, although argumentative tasks were selected for the purpose of this project, they do not fully represent the whole construct of writing and thus the results of this study may not be generalizable to other types of tasks. Further research may produce different results for a wider variety of text types, genres of writing, and modes of production. However, this study was based on the premise that argument tasks have greater potentiality in "encouraging writers to engage in problem solving and, thus, in planning behaviour" (Manchón & Roca de Larios, 2007, p. 563). These tasks are among the requirements of academic writing for many university-level students, and are commonly practiced in

educational contexts. Furthermore, a large number of high-stakes language proficiency tests, such as IELTS or TOEFL, employ such tasks to assess learners' writing proficiency.

Thirdly, the main focus of this study was on the four major metacognitive strategies required for argumentative writing. Yet, this did not mean to downplay the impact of other types of learning strategies within this genre of writing or others. In order to enrich the understanding of L2 writing process and make it more comprehensive, there is a possibility that the scope of the research on written academic genres may go beyond the cognitive/metacognitive domains over more social/cultural/affective underpinnings and domains. After all, writing, as claimed by Cumming (1998), cannot be conducted in a social vacuum and, as rightly argued by Sasaki (2002), the social/cultural contexts within which writing takes place cannot be ignored.

Fourthly, the use of an elaborate six-point scale in the questionnaires seemed to have made it difficult for the participants to pinpoint the gradual differences in the level of strong disagreement to strong agreement. A three-point or five-point scale, as favoured by scholars such as Schellings (2011), and Schellings et al. (2012), could be used to somewhat reduce the variation in the participants' choice of reference points for comparing their conceptions of their performances.

Fifthly, although the contextual influences of the two studies were not intended to be investigated in this project, it is important to acknowledge how second vs. foreign language context or language learning environment could have been a potential and relevant influence, and how the extent of triangulation may have been limited, albeit not eliminated. Therefore, the knowledge gained in this study about the impact of strategy training on the quality of writing may not be easily generalizable or broadly transferrable across FL and SL writing contexts.

The last limitation has to do with maturation and practice effect. In the present study, the changes in participants' knowledge were explored over a six-week period of time in Study 1 and a nine-week period in Study 2 which appeared to have an impact on the development of the specific linguistic traits being studied. Maturation and practice effect factors could lead to changes in the traits attributable to the growth in participants' overall writing knowledge rather than the independent variables.

7.4. Possible future research

The study raises a number of issues that call for further research. Firstly, one may argue that the frequency of an activity does not play a conclusive role in demonstrating the participants' level of awareness. Instead, the time spent on the execution of that particular activity is important. For instance, formulating an action plan prior to the composition may be executed slower or faster for a participant than attending to the formal or structural aspects of a paragraph. Thus, in the future research, it would seem worthwhile to consider the temporal aspects of each strategy as well.

In addition, although the responses to Metacognitive Strategy Questionnaire was compared among four experimental groups and across three time phases, more research is required to explore correlations between the questionnaire instrument and the think-aloud protocols. To do this properly, different study tasks, with different tests, and different age groups may be investigated. This way, the construct-irrelevance variance and the construct underrepresentation, which Messick (1995) sees as major threats to the construct validity of assessments, may vary.

A third area of future research concerns the provision of feedback on participants' writing practices in class. Ideally, a full writing cycle should include the teacher's feedback to help learners to communicate their knowledge, clarify doubts, and engage with their own learning (Wells, 1999). Giving feedback or providing knowledge of the results and evaluative information to learners' written text products can elevate their self-awareness and stimulate reformulation of some wrong concepts they may hold about the nature and structure of genre writing.

Finally, the present study exemplified how the independent variables used in the analyses could help assess the quality of writing and some measures of performance like accuracy and complexity. Further research could investigate some more micro-analytic evidence of the effects of the treatment on the fluency or some other specific linguistic features, as in focused tasks. Practicing such a detailed analysis can help gain a more detailed insight into the linguistic characteristics of argumentative tasks and can provide a more conscientious diagnosis of learners' strengths and weaknesses in their linguistic repertoire. Additionally, a more longitudinal investigation of planning as part of the writing process can be critical

for a comprehensive understanding of L2 learners' writing processes and as an essential contribution to the overall quality of learners' written products.

This study has provided insights into the nature of second language writing in the context of second and foreign language classrooms and its role on L2 leaners' mastery of writing skills in the broader context of L2 writing. In order to demonstrate the relation between planning/metacognitive strategies and learner's writing development, the current study attempted to pinpoint evidence for acquisitional changes in participants' L2 knowledge representations in their writing ability by designing an instruction unit which involved a pre-test, an intervention, and post-test(s). When L2 learners were explicitly instructed metacognitive strategies for writing, significant improvements in quality were shown to have been realised. Teaching specific prewriting, monitoring, and evaluating strategies were shown to have successfully reduced the cognitive overload of the task and allowed participants to formulate higher quality written products. The implications of teaching metacognitive strategies in second language writing education require to be framed as an ongoing action research by educators and as further investigations by other stakeholders on how to nurture the learners' writing development.

REFERENCES

Adams, R., Amani, S., Newton, J., & Alwi, N. A. N. M. (in press). Planning and production in CMC writing In H. Byrnes, & R. M. Manchón (Eds.), *Task-based Language Learning: Insights from and for L2 Writing*. Amsterdam/Philadelphia: John Benjamins.

Afflerbach, P. (2000). Verbal reports and protocol analysis. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (pp. 163-179). Mahwah, NJ: Lawrence Erlbaum Associates.

Ala Amjadi, M. (2012). With the verve of words: Learning foreign languages in Iran. *Tehran Times*. Retrieved from http://www.tehrantimes.com/component/content/article/94791

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge MA: Harvard University Press.

Anderson, N. J. (2002a). The role of metacognition in second language teaching and learning. *ERIC Digest. Education Resources Information Centre*, 56-70.

Anderson, N. J. (2002b). Using telescopes, microscopes, and kaleidoscopes to put metacognition into perspective. *TESOL Matters, 12*(4), 68-72.

Anderson, N. J. (2005). L2 strategy research. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 757-772). Mahwah, NJ: Lawrence Erlbaum Associates

Anderson, N. J. (2007). Metacognition in writing: Facilitating writer awareness. In A. Stubbs, & J. Chapman (Eds.), *Rhetoric, Uncertainty, and the University as Text: How Students Construct the Academic Experience* (pp. 19-43). Regina, Canada: Canadian Plains Research Center, University of Regina.

Anderson, N. J. (2008). Metacognition and good language learners. In C. Griffiths (Ed.), *Lessons from Good Language Learners* (pp. 99-109). Cambridge: Cambridge University Press.

Angouri, J. (2010). Quantitative, qualitative or both? Combining methods in linguistic research. In L. Litosseliti (Ed.), *Research Methods in Linguistics*. UK: Continuum International Publishing Group.

Arsyad, S. (1999). The Indonesian and English argument structure: A cross-cultural rhetoric of argumentative texts. *Australian Review of Applied Linguistics, 22*(2), 85-102.

Atai, M. R., & Mazlum, F. (2012). English language teaching curriculum in Iran: planning and practice. *Curriculum Journal, 24*(3), 389-411.

Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language Testing, 12*(2), 238-257.

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice : Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). The multiple-component model. In A. Miyake, & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). New York: Cambridge University Press.

Bakhtin, M. M. (1981). *The Dialogic Imagination*. Austin: University of Texas Press.
Bakhtin, M. M. (1986). *Speech Genres and Other Late Essays*. Texas: University of Texas Press.
Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York: Freeman.
Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning, 3*(1), 39-58.
Bereiter, C., & Scardamalia, M. (1987). *The Psychology of Written Composition*. Mahwah, NJ: Lawrence Erlbaum Associates.
Berlin, J. A. (1988). Rhetoric and ideology in the writing class. *College English, 50*(5), 477-494.
Biggs, J. B. (1987). *The Study Process Questionnaire (SPQ): Manual*. Hawthorn, Vic.: Australian Council for Educational Research.
Biggs, J. B., & Moore, P. J. (1993). *The Process of Learning*. Sydney, Australia: Prentice-Hall
Bizzell, P. (1986). Composing processes: An overview. In A. Petrosky, & D. Bartholomae (Eds.), *The Teaching of Writing* (pp. 49-70). Chicago: University of Chicago Press.
Boone Jr., H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of Extension, 50*(2). Retrieved from http://www.joe.org/joe/2012april/tt2.php
Borjian, M. (2013). *English in Post-Revolutionary Iran*. Bristol: Multilingual Matters.
Borkowski, J. G. (1992). Metacognitive theory: A framework for teaching literacy, writing, and math skills. *Journal of Learning Disabilities, 25*(4), 253-257.
Bosher, S. (1998). The composing processes of three Southeast Asian writers at the post-secondary level: An exploratory study. *Journal of Second Language Writing, 7*(2), 205-241.
Brown, A. L. (1987). Metacognition, executive control, self regulation and other more mysterious mechanisms. In F. E. Weinert, & R. H. Kluwe (Eds.), *Metacognition, Motivation and Understanding* (pp. 65-116). Hillsdale, NJ: Lawrence Erlbaum.
Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In P. H. Mussen (Ed.), *Handbook of child psychology: Cognitive development* (pp. 77-166). New York: Wiley.
Brown, C., Hedberg, J., & Harper, B. (1994). Metacognition as a basis for learning support software. *Performance Improvement Quarterly, 7*(2), 3-26.
Brown, H. D. (2002). *Strategies for Success: A Practical Guide to Learning English*. White Plains, NY: Pearson Education.
Brown, J. D. (2008). Research methods for applied linguistics: Scope, characteristics, and standards. In A. Davies, & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 476-500). Oxford: Blackwell.
Burtis, P. J., Bereiter, C., Scardamalia, M., & Tetroe, J. (1983). The development of planning in writing. In G. Wells, & B. M. Kroll (Eds.), *Explorations in the Development of Writing* (pp. 153-174). New York: Wiley.
Canagarajah, A. S. (2002). *Critical Academic Writing and Multilingual Students*. Ann Arbor, MI: University of Michigan Press.
Carrell, P. L., Pharis, B. G., & Liberto, J. C. (1989). Metacognitive strategy training for ESL reading. *TESOL Quarterly, 23*(4), 647-678.
Carter, M. (1990). The idea of expertise: An exploration of cognitive and social dimensions of writing. *College Composition and Communication, 41*(3), 265-286.
Chamot, A. U. (2004). Issues in language learning strategy research and teaching. *Electronic Journal of Foreign Language Teaching, 1*(1), 14-26.

Chamot, A. U., & O'Malley, J. M. (1986). *A Cognitive Academic Language Learning Approach: An ESL Content-based Curriculum*. Washington, DC: National Clearinghouse for Bilingual Education.

Chamot, A. U., & O'Malley, J. M. (1994). *The CALLA Handbook: Implementing the Cognitive Academic Language Learning Approach*. Reading, MA: Addison-Wesley Publishing Company.

Clark, H. H., & Clark, E. V. (1977). *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt, Brace, Jovanovich.

Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education, 35*(4), 31-35.

Cohen, A. D. (1994). *Assessing Language Ability in the Classroom*. Boston: Heinle & Heinle.

Cohen, A. D. (1998). *Strategies in Learning and Using a Second Language*. New York: Longman.

Cohen, A. D., & Dörnyei, Z. (2002). Focus on the language learner: Motivation, styles, and strategies. In N. Schmitt (Ed.), *An Introduction to Applied Linguistics* (pp. 170-190). London: Arnold.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education* (5th ed.). New York: RoutledgeFalmer.

Connor, U. (1987). Argumentative patterns in student essays: Cross-cultural differences. In U. Connor, & R. Kaplan (Eds.), *Writing across Languages: Analysis of L2 Text* (pp. 57-71). Reading, MA: Addison Wesley.

Cooper, C. R., & Matsuhashi, A. (1983). A theory of the writing process. In M. Martlew (Ed.), *The Psychology of Written Language* (pp. 3-39). New York: John Wiley.

Costa, A. L. (2008). *The School as a Home for the Mind: Creating Mindful Curriculum, Instruction, and Dialogue*. Thousand Oaks, CA: Corwin Press.

Cottrell, S. (2003). *The Study Skills Handbook*. Basingstoke: Palgrave.

Coutinho, S., Wiemer-Hastings, K., Skowronski, J. J., & Britt, M. A. (2005). Metacognition, need for cognition and use of explanations during ongoing learning and problem solving. *Learning and Individual Differences, 15*(4), 321-337.

Crain-Thoreson, C., Lippman, M. Z., & McClendon-Magnuson, D. (1997). Windows on comprehension: Reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology, 89*(4), 579-591.

Cresswell, A. (2000). Self-monitoring in student writing: Developing learner responsibility. *ELT Journal, 54*(3), 235-244.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Cronin, H., Sinatra, R., & Barkley, W. F. (1992). Combining writing with text organization in content instruction. *NASSP (National Association of Secondary School Principals) Bulletin, 76*(542), 34-45.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*, 367-383.

Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning, 39*, 81-141.

Cumming, A. (1995). Fostering writing expertise in ESL composition instruction: Modeling and evaluation. In D. Belcher, & G. Braine (Eds.), *Academic Writing in a Second Language: Essays on Research and Pedagogy* (pp. 375-397). Norwood, NJ: Ablex.

Cumming, A. (1997). The testing of writing in a second language. In C. Clapham, & D. Corson (Eds.), *Language Testing and Assessment: Encyclopedia of Language and Education* (pp. 51-63). Dordrecht, The Netherlands: Kluwer.

Cumming, A. (1998). Theoretical perspectives on writing. *Annual Review of Applied Linguistics, 18*(1), 61-78.

Cumming, A., & Riazi, A. (2000). Building models of adult second-language writing instruction. *Learning and Instruction, 10*(1), 55-71.

Cumming, A., & So, S. (1996). Tutoring second language text revision: Does the approach to instruction or the language of communication make a difference? *Journal of Second Language Writing, 5*(3), 197-226.

Daller, H., van Hout, R., & Treffers-Daller, J. (2003). Measuring lexical aspects of oral language proficiency among bilinguals: An analysis of different measurements. *Applied Linguistics, 24*(2), 197-222.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of Language Testing*. Cambridge: Cambridge University Press.

De Jong, F. (1987). Differences in the self-regulation processes between successful and less successful students and the prediction of learning performances in case of comprehension and learning of text. In P. R. J. Simons, & G. Beukhof (Eds.), *Regulation of Learning* (pp. 33-45). Den Haag: SVO-Selecta.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* (pp. 121-142). Philadelphia: John Benjamins.

De La Paz, S. (1999). Self-regulated strategy instruction in regular education settings: Improving outcomes for students with and without learning disabilities. *Learning Disabilities Research & Practice, 14*(2), 92-106.

De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology, 97*(2), 139-156.

De La Paz, S. (2007). Managing cognitive demands for writing: Comparing the effects of instructional components in strategy instruction. *Reading & Writing Quarterly, 23*(3), 249-266.

De La Paz, S., & Graham, S. (1997). Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems. *Journal of Educational Psychology, 89*(2), 203.

De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology, 94*(4), 687-698.

DeKeyser, R., Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's "processing instruction: An update". *Language Learning, 52*(4), 805-823.

Delclos, V. R., & Harrington, C. (1991). Effects of strategy monitoring and proactive instruction on children's problem-solving performance. *Journal of Educational Psychology, 83*(1), 35-42.

Dellerman, P., Coirier, P., & Marchand, E. (1996). Planning and expertise in argumentative composition. In G. Rijlaarsdam, v. d. H. Bergh, & M. Couzijn (Eds.), *Theories, Models and Methodology in Writing Research* (pp. 182-195). Amsterdam: Amsterdam University Press.

Denzin, N. K. (1970). *Sociological Methods: A Source Book*. Chicago: Aldine.
Denzin, N. K., & Lincoln, Y. S. (1994). Introduction: Entering the field of qualitative research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 1-17). Thousand Oaks, CA: Sage.
Devine, J. (1993). The role of metacognition in second language reading and writing. In J. G. Carson, & I. Leki (Eds.), *Reading in the Composition Classroom: Second Language Perspectives* (pp. 105-127). Boston: Heinle and Heinle Publishers.
Devine, J., Railey, K., & Boshoff, P. (1993). The implications of cognitive models in L1 and L2 writing. *Journal of Second Language Writing, 2*(3), 203-225.
Dörnyei, Z. (2003). *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Mahwah, NJ: Lawrence Erlbaum Associates.
Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, N.J: Lawrence Erlbaum Associates.
Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. New York: Oxford University Press.
Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 589-630). Oxford: Blackwell.
Dujsik, D. (2008). *The effects of pre-writing strategy training guided by computer-based procedural facilitation on ESL students' strategy use, writing quantity, and writing quality* (PhD Thesis). University of South Florida, United States.
Eckerth, J. (2008). Investigating consciousness-raising tasks: Pedagogically targeted and non-targeted learning gains. *International Journal of Applied Linguistics, 18*(2), 119-145.
Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review, 1*(1), 3-14.
Efklides, A., & Misailidi, P. (2010). Introduction: The present and the future in metacognition. In A. Efklides, & P. Misailidi (Eds.), *Trends and Prospects in Metacognition Research* (pp. 1-18). New York: Springer.
Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9*(1), 1-19.
Ellis, R. (2003). *Task-based Language Teaching and Learning*. Oxford: Oxford University Press.
Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 3-34). Philadelphia: John Benjamins.
Ellis, R. (2008). Explicit form-focused instruction and second language acquisition. In B. Spolsky, & F. M. Hult (Eds.), *The Handbook of Educational Linguistics* (pp. 437-455). Oxford: Blackwell.
Ellis, R. (2012). *Language Teaching Research and Language Pedagogy*. Malden, Mass.: Wiley-Blackwell.
Ellis, R., & Barkhuizen, G. (2005). *Analysing 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*(1), 59-84.
Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 167-192). Philadelphia: John Benjamins.

Emig, J. (1971). *The Composing Processes of Twelfth Graders*. Urbana, IL: National Council of Teachers of English.

Englert, C. S. (1992). Writing Instruction from a sociocultural perspective: The holistic, dialogic, and social enterprise of writing. *Journal of Learning Disabilities, 25*(3), 153-172.

Englert, C. S., Mariage, T. V., & Dunsmore, K. (2006). Tenets of sociocultural theory in writing instruction research. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 208-221). New York: The Guilford Press.

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal, 28*(2), 337-372.

Entwistle, N. J., & Ramsden, P. (1983). Understanding Student Learning. London: Croom Helm.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Eslinger, C. E. (2000). *A more responsive mind: A study of learning strategy adaptation among culturally diverse ESL students* (Unpublished master's thesis). Brigham Young University, Provo, Utah.

Everson, H. T., & Tobias, S. (2001). The ability to estimate knowledge and performance in college: A metacognitive analysis. In H. J. Hartman (Ed.), *Metacognition in Learning and Instruction: Theory, Research, and Practice* (pp. 65-79). Dordrecht, Boston: Kluwer Academic Publishers.

Faigley, L. (1986). Competing theories of process: A critique and proposal. *College English, 48*(6), 527-542.

Field, A. (2005). *Discovering Statistics with SPSS* (2nd ed.). London: SAGE Publications.

Flavell, J. H. (1971). First discussant's comments: What is memory development the development of? *Human Development, 14*(4), 272-278.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231-236). Hillsdale, N.J: Erlbaum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911.

Flavell, J. H., Miller, P. H., & Miller, S. A. (2002). *Cognitive Development* (4 ed.). Upper Saddle River, NJ: Prentice Hall.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail, & J. W. Hagen (Eds.), *Perspectives on the Development of Memory and Cognition*. Hillsdale, NJ: Erlbaum.

Flower, L. (1994). Metacognition: A strategic response to thinking. In L. Flower (Ed.), *The Construction of Negotiated Meaning* (pp. 223-262). Carbondale: Southern Illinois University Press.

Flower, L., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication, 31*(1), 21-32.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365-387.

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*. London: 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.

Fotos, S. S. (1993). Consciousness raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics, 14*(4), 385-407.

Fotos, S. S. (1994). Integrating grammar instruction and communicative language use through grammar consciousness-raising tasks. *TESOL Quarterly, 28*(2), 323-351.

Fotos, S. S., & Ellis, R. (1991). Communicating About Grammar: A Task-Based Approach. *TESOL Quarterly, 25*(4), 605-628.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316-344.

Galbraith, D., & Rijlaarsdam, G. (1999). Effective strategies for the teaching and learning of writing. *Learning and Instruction, 9*(2), 93-108.

Gao, X., & Zhang, L. J. (2011). Joining forces for synergy: Agency and metacognition as interrelated theoretical perspectives on learner autonomy. In G. Murray, X. Gao, & T. Lamb (Eds.), *Identity, Motivation and Autonomy in Language Learning* (pp. 25-41). Bristol: Multilingual Matters Limited.

Garner, R. (1987). *Metacognition and Reading Comprehension*. Norwood, NJ: Ablex Publishing.

Gass, S. (2010). Experimental Research. In B. Paltridge, & A. Phakiti (Eds.), *Continuum Companion to Research Methods in Applied Linguistics* (pp. 7-21). New York: Continuum International Publishing Group.

Ghorbani, M. R. (2009). ELT in Iranian high schools in Iran, Malaysia and Japan: Reflections on how tests influence use of prescribed textbooks. *Reflections on English Language Teaching, 8*(2), 131-139.

Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. Amsterdam/Philadelphia: John Benjamins.

Gilbert, K. (2004). A comparison of argument structures in L1 and L2 student writing. *Journal of Asian Pacific Communication, 14*(1), 55-76.

Goh, C. (1997). Metacognitive awareness and second language listeners. *ELT Journal, 51*(4), 361-369.

Goodrich Andrade, H. G., & Perkins, D. N. (1998). Learnable Intelligence and Intelligent Learning. In R. J. Sternberg, & W. M. Williams (Eds.), *Intelligence, instruction and assessment: Theory in practice* (pp. 67-94). Mahwah, NJ: Lawrence Erlbaum Associates.

Gosden, H. (1996). Verbal reports of Japanese novices' research writing practices in English. *Journal of Second Language Writing, 5*(2), 109-128.

Gould, J. D. (1980). Experiments on composing letters: Some facts, some myths, and some observations. In W. Gregg, & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 97-127). Hillsdale, NJ: Erlbaum.

Gourgey, A. F. (2001). Metacognition in basic skills instruction. In H. J. Hartman (Ed.), *Metacognition in Learning and Instruction: Theory, Research and Practice* (pp. 17-32). Dordrecht, Boston: Kluwer Academic Publishers.

Grabe, W., & Kaplan, R. B. (1996). *Theory and Practice of Writing: An Applied Linguistic Perspective*. New York: Longman.

Graham, S. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology, 82*(4), 781.

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 187-207). New York: Guilford.

Graham, S., & Harris, K. R. (1993). Self-regulated strategy development: Helping students with learning problems develop as writers. *Elementary School Journal, 94*(2), 169-181.

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3-12.

Graham, S., & Harris, K. R. (2007). Best practices in teaching planning. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (pp. 119-140). New York: Guilford Press.

Graham, S., & Perin, D. (2006). *Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High Schools*. Washinton, DC: Alliance for Excellent Education.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476.

Graham, S., Schwartz, S. S., & MacArthur, C. A. (1993). Knowledge of writing and the composing process, attitude toward writing, and self-efficacy for students with and without learning disabilities. *Journal of Learning Disabilities, 26*(4), 237-249.

Graves, K. (1996). A framework of course development processes. In K. Graves (Ed.), *Teachers as Course Developers*. Cambridge: Cambridge University Press.

Grenfell, M., & Harris, V. (1999). *Modern Languages and Learning Strategies: In Theory and Practice*. London: Routledge.

Griffiths, C. (2008). Strategies and good language learners. In C. Griffiths (Ed.), *Lessons from Good Language Learners* (pp. 83-98). Cambridge: Cambridge University Press.

Gu, P. Y. (1996). Robin Hood in SLA: What has the learning strategy researcher taught us? *Asian Journal of English Language Teaching, 6*, 1-29.

Hacker, D. J., Keener, M. C., & Kircher, J. C. (2009). Writing is applied metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 154-172). New York: Routledge.

Haghverdi, H. R., Biria, R., & Khalaji, H. R. (2013). The impact of task-planning and gender on the accuracy of narrations composed by Iranian EFL learners. *Journal of Language Teaching and Research, 4*(1), 74-83.

Halliday, M. A. K. (1978). *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Hamp-Lyons, L. (1995). Rating nonnative writing: The trouble with holistic scoring. *TESOL Quarterly, 29*(4), 759-762.

Harris, K. R., & Graham, S. (1992). Self-regulated strategy development: A part of the writing process. In M. Pressley, K. R. Harris, & J. T. Guthrie (Eds.), *Promoting Academic Competence and Literacy in School* (pp. 277-309). New York: Academic Press.

Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and children's writing. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 131-153). New York: Routledge.

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43*(2), 295-340.

Harris, K. R., Graham, S., & Pressley, M. (1992). Cognitive-behavioral approaches in reading and written language: Developing self-regulated learners. In N. N. Singh, & I. L. Beale (Eds.), *Learning Disabilities: Nature, Theory, and Treatment* (pp. 415-451). New York: Springer-Verlag.

Harris, K. R., Santangelo, T., & Graham, S. (2010). Metacognition and strategies instruction in writing. In H. S. Waters, & W. Schneider (Eds.), *Metacognition, Strategy Use, and Instruction* (pp. 226-256). New York: Guilford Press.

Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.), *Metacognition in Learning and Instruction: Theory, Research, and Practice* (pp. 33-68). Dordrecht, Boston: Kluwer Academic Publishers.

Hartman, H. J., & Sternberg, R. J. (1993). A broad BACEIS for improving thinking. *Instructional Science, 21* (5), 401-425.

Hawthorne, S. (2008). *Engaging reluctant writers: The nature of reluctance to write and the effect of a self-regulation strategy training programme on the engagement and writing performance of reluctant writers in secondary school English* (PhD Thesis). University of Auckland, New Zealand.

Hayati, A. M., & Mashhadi, A. (2010). Language planning and language-in-education policy in Iran. *Language Problems & Language Planning, 34*(1), 24-42.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. L. S. Ransdell (Ed.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications* (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum Associates.

Hayes, J. R., & Flower, L. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 3-30). Hillsdale, NJ: Erlbaum.

Hayes, J. R., & Flower, L. (1983). Uncovering cognitive processes in writing: An introduction to protocol analysis. In P. mosenthal, L. Tamor, & S. Walmsley (Eds.), *Research in Writing: Principles and Methods*. New York: Longman.

Hayes, J. R., & Flower, L. (1986). Writing research and the writer. *American Psychologist, 41*(10), 1106-1113.

Hayes, J. R., & Nash, J. G. (1996). On the nature of planning in writing. In M. Levy, & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications* (pp. 29-55). Mahwah, NJ: Lawrence Erlbaum Associates.

He, T. H. (2005). Effects of mastery and performance goals on the composition strategy use of adult EFL writers. *Canadian Modern Language Review, 61*(3), 407-431.

Hill, J. R., & Hannofin, M. J. (1997). Cognitive Strategies and Learning from the World Wide Web. *Educational Technology Research and Development, 45*(4), 37-64.

Hosseini, S. M. H. (2007). ELT in Higher Education in Iran and India: A critical view. *Language in India, 7*, 1-10.

Hunt, K. (1965). *Grammatical structures written at three grade levels. NCTE research report No. 3.* . Champaign, IL: NCTE.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics, 34*(8), 1091-1112.

Hyland, K. (2003). *Second Language Writing*. Cambridge: Cambridge University Press.

Hyland, K. (2010). Researching writing. In B. Paltridge, & A. Phakiti (Eds.), *Continuum Companion to Research Methods in Applied Linguistics* (pp. 191-204). New York: Continuum International Publishing Group.

Hyland, K. (2013). Writing in the university: Education, knowledge and reputation. *Language Teaching, 46*(1), 53-70.

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist, 22*(3-4), 255-278.

Janssen, D., Van Waes, L., & Van den Bergh, H. (1996). Effects of thinking aloud on writing processes. In C. M. Levy, & R. S (Eds.), *Science of Writing: Theories, Methods, Individual Differences, and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Johns, A. M. (2003). Genre and ESL/EFL composition instruction. In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing* (pp. 195-217). New York: Cambridge University Press.

Johns, A. M. (Ed.). (2002). *Genre in the Classroom: Multiple Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson, D. M., & Roen, D. H. (Eds.). (1989). *Richness in Writing: Empowering ESL Students*. New York: Longman.

Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing, 21*(3), 264-282.

Johnson, R. K. (1989). A decision-making framework for the coherent language curriculum. In R. K. Johnson (Ed.), *The Second Language Curriculum* (pp. 1-23). Cambridge: Cambridge University Press.

Jong, Y. O. (2009). *An investigation into the benefits of collaborative writing for the development of EFL children's communication skills: A reflective report of a teacher researcher* (PhD Thesis). University of Warwick, UK.

Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. *Language Learning, 16*(1-2), 1-20.

Karim, H. A. (2010). *The use of L1 and L2 in prewriting discussions in EFL writing and students 'attitudes towards L1 and L2 use in prewriting discussions* (Master's Dissertation). Bilkent University, Turkey.

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition, 15*(3), 256-266.

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. *The American Journal of Psychology, 103*(3), 327-342.

Kellogg, R. T. (1994). *The Psychology of Writing*. New York: Oxford University Press.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy, & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications* (pp. 57-71). Mahwah, NJ: Lawrence Erlbaum Associates.

Khomeijani Farahani, A. A., & Meraji, S. R. (2011). Cognitive task complexity and L2 narrative writing performance. *Journal of Language Teaching and Research, 2*(2), 445-456.

Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2006). Writing as a learning tool: Testing the role of students' writing strategies. *European Journal of Psychology of Education, 21*(1), 17-34.

Kieft, M., Rijlaarsdam, G., & van den Bergh, H. (2008). An aptitude–treatment interaction approach to writing-to-learn. *Learning and Instruction, 18*(4), 379-390.

Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction, 5*(2), 143-165.

Kluwe, R. H. (1987). Executive decisions and regulation of problem solving. In F. Weinert, & R. Kluwe (Eds.), *Metacognition, Motivation, and Understanding* (pp. 31-64). Hillsdale, NJ: Lawrence Erlbaum.

Kolencik, P. L., & Hillwig, S. A. (2011). *Enhancing Metacognition: Supporting Learners through Metacognitive Teaching Strategies*. New York: P. Lang.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing, 20*(2), 148-161.

Kramer-Dahl, A., & Chia, A. (2012). Enriching the space of learning through weaving: How a teacher builds her students' understanding of (a) genre over time. *Pedagogies: An International Journal, 7*(1), 72-94.

Krause, K.-L., & O'Brien, D. (1999). A sociolinguistic study of the argumentative writing of Chinese students. *Education Journal, 27*(2), 43-63.

Kroll, B. (1978). Sorting out writing problems. In C. H. Blatchfold, & J. Schachter (Eds.), *TESOL '78: EFL Policies, Programs, Practices* (pp. 176–182). Washington, DC: TESOL.

Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second Language Writing: Research Insights for the Classroom* (pp. 140-154). Cambridge Cambridge University Press.

Kroll, B. (2003). Introduction: Teaching the next generation of second language writers. In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing* (pp. 1-10). New York: Cambridge University Press.

Ku, K., & Ho, I. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning, 5*(3), 251-267.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*(4), 285-327.

Lally, C. G. (2000). First language influences in second language composition: The effect of pre-Writing. *Foreign Language Annals, 33*(4), 428-432.

Lane, K. L., Harris, K. R., Graham, S., Weisenbach, J. L., Brindle, M., & Morphy, P. (2008). The effects of self-regulated strategy development on the writing performance of second-grade students with behavioral and writing difficulties. *The Journal of Special Education, 41*(4), 234-253.

Lansman, M., Smith, J. B., & Weber, I. (1990). *Using computer-generated protocols to study writers' planning strategies* (Technical Report 90-033). Chapel Hill: University of North Carolina, Department of Computer Science.

Lantolf, J. P. (Ed.). (2000). *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.

Larkin, S. (2009). Socially mediated metacognition and learning to write. *Thinking Skills and Creativity, 4*(3), 149-159.

Leki, I. (2007). *Undergraduates in a Second Language: Challenges and Complexities of Academic Literacy Development*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Lin, B. (2006). Genre-based teaching and Vygotskian principles in EFL: The case of a university writing course. *Asian EFL journal, 8*(3), 267-280.

Lin, S. C., Monroe, B. W., & Troia, G. A. (2007). Development of writing knowledge in grades 2-8: A comparison of typically developing writers and their struggling peers. *Reading & Writing Quarterly, 23*(3), 207-230.

Livingston, J. A. (1997). Metacognition: An overview. *Retrieved May 2013 from http://gse.buffalo.edu/fas/shuell/cep564/Metacog.htm*.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. D. Lambert, & E. Shohamy (Eds.), *Language Policy and Pedagogy: Essays in Honor of A. Ronald Walton* (pp. 179-192). Amsterdam: John Benjamins.

Lv, F., & Chen, H. (2010). A study of metacognitive-strategies-based writing instruction for vocational college students. *English Language Teaching, 3*(3), 136-144.

Macaro, E. (2001). *Learning Strategies in Foreign and Second Language Classrooms*. London: Continuum.

MacArthur, C. A., & Graham, S. (1987). Learning disabled students' composing under three methods of text production handwriting, word processing, and dictation. *The Journal of Special Education, 21*(3), 22-42.

MacArthur, C. A., Graham, S., Schwartz, S. S., & Schafer, W. D. (1995). Evaluation of a writing instruction model that integrated a process approach, strategy instruction, and word processing. *Learning Disability Quarterly, 18*(4), 278-291.

Mahnam, L., & Nejadansari, D. (2012). The effects of different pre-writing strategies on Iranian EFL writing achievement. *International Education Studies, 5*(1), 154-160.

Manchón, R. M. (2009). Introduction: Broadening the perspective of L2 writing scholarship: The contribution of research on foreign language writing. In R. M. Manchón (Ed.), *Writing in Foreign Language Contexts: Learning, Teaching, and Research* (pp. 1-19). Bristol: Multilingual Matters.

Manchón, R. M., Murphy, L., & Roca de Larios, J. (2005). Using concurrent protocols to explore L2 writing processes: Methodological issues in the collection and analysis of data. In P. K. Matsuda, & T. J. Silva (Eds.), *Second Language Writing Research: Perspectives on the Process of Knowledge Construction* (pp. 191-205). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Manchón, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing. *Language Learning, 57*(4), 549-593.

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2000). An approximation to the study of backtracking in L2 writing. *Learning and Instruction, 10*(1), 13-35.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2002). *Classroom Instruction that Works: Research-based Strategies for Increasing Student Achievement*. Alexandria, VA.: Association for Supervision and Curriculum Development.

Matsuda, P. K. (2003a). Process and post-process: A discursive history. *Journal of Second Language Writing, 12*(1), 65-83.

Matsuda, P. K. (2003b). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing* (pp. 15-34). New York: Cambridge University Press.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*(1-2), 49-63.

Mayer, R. E. (2001). What good is educational psychology? The case of cognition and instruction. *Educational Psychologist, 36*(2), 83-88.

McCormick, C. B. (2003). Metacognition and Learning. In W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of Psychology* (pp. 79-102). Hoboken, NJ: John Wiley & Sons.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*(3), 299-325.

McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 115-130). New York: Guilford.

McNamara, T. F. (1996). *Measuring Second Language Performance*. London: Longman.
Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*(1), 83-108.
Meijer, J., Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation, 12*(3), 209-237.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.
Moragne e Silva, M. (1989). A study of composing in a first and second language. *Texas Papers in Foreign Language Education, 1*(2), 132-151.
Muis, K. R., Winne, P. H., & Jamieson-Noel, D. (2007). Using a multitrait-multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *British Journal of Educational Psychology, 77*(1), 177-195.
Murray, D. M. (1972). Teach writing as a process not product. *The Leaflet, 71*(3), 11-14.
Murray, N. (2010). Considerations in the post-enrolment assessment of English language proficiency: Reflections from the Australian context. *Language Assessment Quarterly, 7*(4), 343-358.
Naiman, N., Frohlich, M., Stern, H., & Todesco, A. (1978). *The good language learner. Research in Education Series No.7*. Toronto: The Ontario Institute for Studies in Education.
Nation, I. S. P., & Macalister, J. (2010). *Language Curriculum Design*. New York: Routledge.
Nelson, G. L., & Carson, J. G. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing, 7*(2), 113-131.
Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
Nguyen, L. T. C., & Gu, Y. (2013). Strategy-based instruction: A learner-focused approach to developing learner autonomy. *Language Teaching Research, 17*(1), 9-30.
Nunan, D. (1996). Learner strategy training in the classroom: An action research study. *TESOL Journal, 6*(1), 35-41.
Nystrand, M. (1989). A social-interactive model of writing. *Written Communication, 6*(1), 66-85.
Nystrand, M. (Ed.). (1982). *What Writers Know: The Language, Process and Structure of Written Discourse*. New York: Academic Press.
O'Malley, J. M., & Chamot, A. U. (1990). *Language Strategies in Second Language Acquisition*. New York: Cambridge University Press.
O'Neil Jr, H. F., & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *The Journal of Educational Research, 89*(4), 234-245.
Ojima, M. (2006). Concept mapping as pre-task planning: A case study of three Japanese ESL writers. *System, 34*(4), 566-585.
Olive, T., Piolat, A., & Roussey, J. Y. (1997). *The effect of the degree of expertise and of the level of knowledge on cognitive effort associated to writing processes.* Paper presented at the European Writng Conferences, Barcelona.
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*(4), 218-233.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21*(1), 109-148.

Ortega, L. (2009). Studying writing across EFL contexts: Looking back and moving forward. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 232-255). Bristol: Multilingual Matters.

Oxford, R. (1990). *Language Learning Strategies: What Every Teacher Should Know*. Boston: Heinle & Heinle.

Oxford, R. (1994). *Language learning strategies: An update*. Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED376707).

Oxford, R. (2001). *Integrated Skills in the ESL/EFL Classroom*. Washington, DC: ERIC Clearinghouse on Languages and Linguistics. (ERIC Reproduction Service No. ED 456670).

Oxford, R., & Crookall, D. (1989). Research on language learning strategies: Methods, findings, and instructional issues. *The Modern Language Journal, 73*(4), 404-419.

Oxford, R., Crookall, D., Cohen, A. D., Lavine, R., Nyikos, M., & Sutter, W. (1990). Strategy training for language learners: Six situational case studies and a training model. *Foreign Language Annals, 23*(3), 197-216.

Oxford, R., & Leaver, B. L. (1996). A synthesis of strategy instruction for language learners. In R. L. Oxford (Ed.), *Language Learning Strategies Around the World: Cross-cultural Perspectives* (pp. 227-246). Honolulu, HI: University of Hawaii Press.

Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology, 91*(2), 230.

Paltridge, B. (2002). Genre, text type, and the English for Academic Purposes (EAP) classroom. In A. M. Johns (Ed.), *Genre in the Classroom: Multiple Perspectives* (pp. 69-88). Mahwah, NJ: Lawrence Erlbaum Associates.

Pieronek, F. (1994). Using maps to teach note taking and outlining for report writing. *The Social Studies, 85*(4), 165-169.

Pincas, A. (1962). Structural linguistics and systematic composition teaching to students of English as a foreign language. *Language Learning, 12*(3), 185-194.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice, 41*(4), 219-225.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Piolat, A., Roussey, Y. J., & Rous, P. (1996). *Effects of attention directed to different texts on the mental effort related to writing processes.* Paper presented at the European Writing Conference, Barcelona.

Piri, F., Barati, H., & Ketabi, S. (2012). The effects of pre-task, on-line, and both pre-task and on-line planning on fluency, complexity, and accuracy – The case of Iranian EFL learners' written production. *English Language Teaching, 5*(6), 158-167.

Pressley, M., & McCormick, C. (1995). *Cognition, Teaching, and Assessment*. New York: Harper Collins College Publishers.

Prins, F. J., Veenman, M. V. J., & Elshout, J. J. (2006). The impact of intellectual ability and metacognition on learning: New support for the threshold of problematicity theory. *Learning and Instruction, 16*(4), 374-387.

Prior, P. (2006). A sociocultural theory of writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 54-66). New York: Guilford.

Pu, X. (2009). Effects of prewriting discussion on the language quality of argumentative writings. *CELEA Journal, 32*(5), 16-24.

Purpura, J. E. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning, 47*(2), 289-325.

Purpura, J. E. (1999). *Learner Strategy Use and Performance on Language Tests: A Structural Equation Modeling Approach*. Cambridge: Cambridge University Press.

Rahimpour, M., & Nariman-Jahan, R. (2011). The effects of planning on writing narrative task performance with low and high EFL proficiency. *English Language Teaching, 4*(1), 120-127.

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.

Raimes, A. (1983). Tradition and revolution in ESL teaching. *TESOL Quarterly, 17*(4), 535-552.

Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly, 19*(2), 229-258.

Ramanathan, V. (2003). Written textual production and consumption (WTPC) in vernacular and English-medium settings in Gujarat, India. *Journal of Second Language Writing, 12*(2), 125-150.

Ransdell, S. (1995). Generating thinking-aloud protocols: Impact on the narrative writing of college students. *The American Journal of Psychology, 108*(1), 89-98.

Ransdell, S., Lavelle, B., & Levy, C. M. (2002). The effects of training a good working memory strategy on L1 and L2 writing. In S. Ransdell, & M. L. Barbiere (Eds.), *New Directions for Research in L2 Writing* (pp. 133-144). Dordrecht: Kluwer Academic Publishers.

Rees-Miller, J. (1993). A critical appraisal of learner training: Theoretical bases and teaching implications. *TESOL Quarterly, 27*(4), 679-689.

Reichelt, M. (2003). Defining "good writing": A cross-cultural perspective. *Composition Studies, 31*(1), 99-126.

Riazi, A. M., & Mosalanejad, N. (2010). Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using Bloom's taxonomy. *TESL-EJ: The Electronic Journal for English as a Second Language, 13*(4), 1-16.

Rijlaarsdam, G., & Van den Bergh, H. (1996). The dynamics of composing—An agenda for research into an interactive compensatory model of writing: Many questions, some answers. In C. M. Levy, & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences and Applications* (pp. 107-125). Mahwah, NJ: Lawrence Erlbaum Associates.

Roberts, K. (2012). Education at a Glance: OECD Indicators 2012: Country note: Australia. Retrieved February 2013, from http://www.oecd.org/edu/EAG2012%20-%20Country%20note%20-%20Australia.pdf

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning, 45*(1), 99-140.

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287-318). Cambridge: Cambridge University Press.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics, 22*(1), 27-57.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*(2), 45-105.

Roca de Larios, J. (1996). Linearization strategies in EFL writing: Some observations. *Lenguaje y Textos, 8*, 191-208.

Roca de Larios, J., Murphy, L., & Manchón, R. M. (1999). The use of restructuring strategies in EFL writing: A study of Spanish learners of English as a foreign language. *Journal of Second Language Writing, 8*(1), 13-44.

Rogers, C. R. (1970). Communication: Its blocking and facilitation. In R. Young, A. Becker, & K. Pike (Eds.), *Rhetoric: Discovery and Change* (pp. 284-288). New York: Harcourt.

Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly, 9*(1), 41-51.

Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal, 38*(2), 437-460.

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology, 77*(2), 351-378.

Sánchez-Alonso, S., & Vovides, Y. (2007). Integration of metacognitive skills in the design of learning objects. *Computers in Human Behavior, 23*(6), 2585-2595.

Sanguran, 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). Philadelphia: John Benjamins.

Sasaki, M. (2002). Building an empirically-based model of EFL learners' writing processes. In S. Ransdell, & M.-L. Barbier (Eds.), *New Directions for Research in L2 Writing* (pp. 49-80). Dordrecht: Kluwer Academic Publishers.

Sasaki, M. (2004). A Multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning, 54*(3), 525-582.

Schellings, G. L. M. (2011). Applying learning strategy questionnaires: Problems and possibilities. *Metacognition and Learning, 6*(2), 91-109.

Schellings, G. L. M., van Hout-Wolters, B. H. A. M., Veenman, M. V. J., & Meijer, J. (2012). Assessing metacognitive activities: The in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education, 28*(3), 963-990.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. W. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in Second Language Learning, AILA Review 11*, 237-326.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-33). Cambridge: Cambridge University Press.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*(1-2), 113-125.

Schraw, G. (2001). Promoting general metacognitive awareness. In H. J. Hartman (Ed.), *Metacognition in Learning and Instruction: Theory, Research and Practice* (pp. 3-16). Dordrecht, Boston: Kluwer Academic Publishers.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460-475.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351-371.

Schultz, J. M. (1991). Mapping and cognitive development in the teaching of foreign language writing. *The French Review, 64*(6), 978-988.

Scruggs, T. E., Mastropieri, M. A., Monson, J., & Jorgensen, C. (1985). Maximizing what gifted students can learn: Recent findings of learning strategy research. *Gifted Child Quarterly, 29*(4), 181-185.

Sengupta, S. (2000). An investigation into the effects of revision strategy instruction on L2 secondary school learners. *System, 28*(1), 97-113.

Sexton, M., Harris, K. R., & Graham, S. (1998). Self-regulated strategy development and the writing process: Effects on essay writing and attributions. *Exceptional Children, 64*(3), 295-311.

Shaughnessy, M. P. (1976). Diving in: An introduction to basic writing. *College Composition and Communication, 27*(3), 234-239.

Shaw, S. D. (2004). IELTS writing: revising assessment criteria and scales (Phase 3). *Research Notes, 16*, 3-7.

Shaw, S. D., & Weir, C. J. (2007). *Examining Writing: Research and Practice in Assessing Second Language Writing* (Vol. 26): Cambridge University Press.

Shen, H. J. (2003). The role of explicit instruction in ESL/EFL reading. *Foreign Language Annals, 36*(3), 424-433.

Shi, L. (1998). Effects of prewriting discussions on adult ESL students' compositions. *Journal of Second Language Writing, 7*(3), 319-345.

Shiau, Y.-S. S., & Adams, R. (2011). The effects of increasing reasoning demands on accuracy and complexity in L2 oral production. *University of Sydney Papers in TESOL, 6*, 121-146.

Shin, Y. (2008). *The effects of planning on L2 writing: A study of Korean learners of English as a foreign language* (PhD Thesis). University of Iowa, United States.

Shohamy, E. (1983). The stability of the oral proficiency trait on the oral interview speaking test. *Language Learning, 33*(4), 527-539.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27*(4), 657-677.

Silva, T., Leki, I., & Carson, J. (1997). Broadening the perspective of mainstream composition studies: Some thoughts from the disciplinary margins. *Written Communication, 14*(3), 398-428.

Sitko, B. M. (1998). Knowing how to write: Metacognition and writing instruction. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 93-115). Mahwah, NJ: Lawrence Erlbaum Associates.

Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis, & D. Willis (Eds.), *The Challenge and Change in Language Teaching* (pp. 17-30). Oxford: Heinemann.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185-211.

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. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 183-205). Cambridge: Cambridge University Press.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioural and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493-525.

Smagorinsky, P. (1991). The writer's knowledge and the writing process: A protocol analysis. *Research in the Teaching of English, 25*(3), 339-364.

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation, 10*(2), 117-139.

Stapa, S. H., & Abdul Majid, A. (2009). The use of first language in developing ideas in second language writing. *European Journal of Social Sciences, 7*(4), 41-47.

Stern, H. H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review, 31*(4), 304-318.

Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols: Issues for research. In P. Smagorinsky (Ed.), *Speaking about Writing: Reflections on Research Methodology* (pp. 89-111). Thousand Oaks, CA: Sage Publications.

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 99-118). Harlow, England: Longman.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Los Angeles: Sage Publications.

Thomas, G. P. (2003). Conceptualisation, development and validation of an instrument for investigating the metacognitive orientation of science classroom learning environments: The Metacognitive Orientation Learning Environment Scale-Science (MOLES-S). *Learning Environments Research, 6*(2), 175-197.

Thomas, G. P., Anderson, D., & Nashon, S. (2008). Development of an instrument designed to investigate elements of science students' metacognition, self-efficacy and learning processes: The SEMLI-S. *International Journal of Science Education, 30*(13), 1701-1724.

Tillema, M., van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2011). Relating self reports of writing behaviour and online task execution using a temporal model. *Metacognition and Learning, 6*(3), 229-253.

Troia, G. A., & Graham, S. (2002). The effectiveness of a highly explicit, teacher-directed strategy instruction routine changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities, 35*(4), 290-305.

Tsai, H. L. (2004). *Investigating the relationships between ESL writers' strategy use and their second language writing ability* (PhD Thesis). Teachers College, Columbia University

Van Den Bergh, H., & Rijlaarsdam, G. (1996). The dynamics of composing: Modeling writing process data. In C. M. Levy, & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications* (pp. 207-232). Mahwah, NJ: Lawrence Erlbaum Associates.

Van Den Bergh, H., & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process and relationships with text quality. *Educational Psychology, 21*(4), 373-385.

Van Hout-Wolters, B. (2009). Measuring Learning strategies: Different kinds of assessment methods and their usefulness in education and research. *Pedagogische Studiën, 86*, 110-103.

Van Hout-Wolters, B., & Duffy, T. (2000). Assessing active self-directed learning. In R. Simons, & J. v. d. Linden (Eds.), *New Learning* (pp. 83-101). Dordrecht: Kluwer.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. London: Academic Press.

Vance, S. J. (1999). *Language learning strategies: Is there a best way to teach them?* Unpublished manuscript. (ERIC Document Reproduction Service ED 438-716; FL 026-146).

Vandergrift, L. (2002). "It was nice to see that our predictions were right": Developing metacognition in L2 listening comprehension. *Canadian Modern Language Review, 58*(4), 555-575.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and Research* New York: Ablex.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning, 52*(4), 755-803.

Veenman, M. V. J. (1993). *Intellectual ability and metacognitive skill: Determinants of discovery learning in computerized learning environments* (PhD Thesis). University of Amsterdam, The Netherlands.

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt, & B. Moschner (Eds.), *Learning Strategies and Metacognition: Implications for Research and Practice*. Berlin, Germany: Waxmann.

Veenman, M. V. J., & Beishuizen, J. J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. *Learning and Instruction, 14*(6), 621-640.

Veenman, M. V. J., Elshout, J. J., & Groen, M. G. M. (1993). Thinking aloud: Does it affect regulatory processes in learning? *Tijdschrift voor Onderwijsresearch, 18*(6), 322-330.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15*(2), 159-176.

Veenman, M. V. J., Van Hout-Wolters, B., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3-14.

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*(1), 89-109.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65-83.

Washington, V. M. (1988). Report writing: A practical application of semantic mapping. *Teacher Educator, 24*(1), 24-30.

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction, 9*(6), 493-516.

Weigle, S. C. (2002). *Assessing Writing*. New York: Cambridge University Press.
Weinert, F. E. (1987). Introduction and overview: Metacognition and motivation as determinants of effective learning and understanding. In F. E. Weinert, & R. H. Kluwe (Eds.), *Metacognition, Motivation, and Understanding* (pp. 1-16). Hillsdale, NJ: Lawrence Erlbaum.
Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 315-327). New York: Macmillan.
Weinstein, C. E., Schuke, A. C., & Palmer, D. P. (1987). *Learning and Study Strategies Inventory*. Clearwater, FL: H & H Publishing.
Wells, G. (1999). *Dialogic Inquiry: Towards a Sociocultural Practice and Theory of Education*. Cambridge: Cambridge University Press.
Wendel, J. N. (1997). *Planning and second-language narrative production* (PhD Thesis). Temple University, Japan.
Wenden, A. L. (2002). Learner development in language learning. *Applied Linguistics, 23*(1), 32-55.
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.
Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*(1), 85-106.
Wikipedia. (2009). Communications in Iran. Retrieved August 2012, from http://en.wikipedia.org/wiki/Communications_in_Iran
Xiao, Y. (2007). Applying metacognition in EFL writing instruction in China. *Reflections on English Language Teaching, 6*(1), 19-33.
Young, R. (1978). Paradigms and problems: Needed research in rhetorical invention. In C. R. Cooper, & L. Odel (Eds.), *Research on Composing: Points of Departure* (pp. 29-47). Urbana, IL: National Council of Teachers of English.
Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning, 45*(1), 3-42.
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.
Zamel, V. (1976). Teaching composition in the ESL classroom: What we can learn from the research in the teaching of English. *TESOL Quarterly, 10*(1), 67-76.
Zhang, L. J. (2001). Awareness in reading: EFL students' metacognitive knowledge of reading strategies in an acquisition-poor environment. *Language Awareness, 10*(4), 268-288.
Zhang, Y. (2010). Investigating the role of metacognitive knowledge in English Writing. *Papers in Applied Language Studies, 14*, 25-46.
Zhigang, M., & Xudong, W. (2008). Interpreting performance under different task-planning conditions. *Babel, 54*(3), 201-233.
Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written Communication, 12*(4), 492-528.
Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*(1), 51.
Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22*(1), 73-101.

APPENDIX A. Metacognitive Strategy Questionnaire (MSQ)

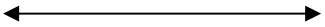
Gender: F M
Given Name: ……
Age: ……
How long have you studied English as a foreign or second language? ……

Instructions: Please read each statement and show how often you used the strategy in the writing task you just had by checking the appropriate box:

1 means "**Disagree**", while **6** indicates "**Agree**".

If you have any questions about any of the items, please raise your hands.

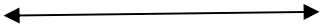
Disagree *Agree*



Before I started writing,	**1**	**2**	**3**	**4**	**5**	**6**
1. I planned for the content of my essay.						
2. I planned how I was going to structure different parts of the essay.						
3. I planned what language features I was going to use in my essay.						
4. I thought about how much time I should spend on the essay.						
5. I thought about the length of my essay.						

When I wrote, I tried to	**1**	**2**	**3**	**4**	**5**	**6**
6. think about what questions my reader might ask about the topic.						
7. think about my reader's views and opinions on the topic.						
8. choose the level of formality of my essay to suit the reader.						

Disagree *Agree*



When I was writing, I tried to think about	**1**	**2**	**3**	**4**	**5**	**6**
9. how much time I had remaining.						
10. whether I was spelling some words correctly.						
11. whether I was using appropriate vocabulary.						
12. whether I was using the correct grammar (e.g. tenses, prepositions, etc.)						
13. how many arguments I should have in the essay.						
14. whether the arguments followed the instruction of the essay.						
15. what parts my essay should have.						
16. how to connect different parts of my essay.						

After writing, I reread my essay and	**1**	**2**	**3**	**4**	**5**	**6**
17. made sure the language of my essay was clear.						
18. made sure the organisation was easy to follow.						
19. made sure I had covered the content fully.						
20. made sure all the paragraphs were relevant to the topic.						

APPENDIX B. Sample lesson plan

Session 1	**Groups +M+P & +M-P**
Aim	Metacognitive strategy: Planning Writing element: Introduction paragraph
Learning Outcomes	By the end of this session, participants will: - Be aware of the significance of planning strategy in writing - Practice planning strategy in writing an essay - Learn various strategies for generating ideas about a topic - Learn how to develop introduction paragraph
Stages	**1. Introducing learning outcomes of the course** (3 min) **2. Understanding the topic** (5 min) **3. Introducing rhetorical modes of discourse in essay writing** (2 min) **4. Overview of writing essay parts** (5 min) **5. Lead-in: Life without a plan!** (5 min) **6. Introducing** "**Planning" in writing** (5 min) **7. Planning (Time)** (5 min) **8. Planning (Content)** (5 min) **9. Planning (Language Features)** (15 min) **10. Introduction paragraph** (15 min) **11. Break time + distributing writing sheets** (10 min) **12. Writing Practice** (45 min)

Stage 1:
Aim: Introducing learning outcomes of the course
Time: 3 minutes
Procedure:

1. Researcher thanks the participants for their participation in the course and shows the first PowerPoint slide which contains the following quotation:
   - *"Most people won't realize that writing is a craft. You have to take your apprenticeship in it like anything else."* (Katherine Anne Porter)
2. Researcher asks participants what an essay is and then summarises the answers by saying that an essay is a critical evaluation which communicates an assemblage of facts and opinions about a certain topic.

Stage 2:
Aim: Understanding the topic
Time: 5 minutes
Procedure:

1. Researcher explains that it is essential to read the topic carefully and take particular note of the instruction given.
2. Researcher then shows the next slide containing a list of some of the verbs which are often used in essay questions.

Account for: give the reasons for.
Analyse: examine in depth and describe the main characteristics of.
Assess: weigh up the elements of and arrive at a conclusion about.
Comment: give an opinion and provide evidence for your views.
Compare: bring out the similarities between.
Contrast: bring out the dissimilarities between.
Define: explain the exact meaning of.
Describe: use words and diagrams to illustrate.
Discuss: provide evidence or opinions about; arriving at a balanced conclusion.
Evaluate: weigh up or appraise.
Explain: make the meaning of something clear.
Illustrate: use diagrams or examples to make clear.
Justify: show that an idea or statement is correct.
List: provide an itemised series of statements about.
Outline: describe the essential parts only.
Review: examine critically.
State: express clearly.
Summarise: without illustrations, provide a brief account of.

Stage 3:
Aim: Introducing rhetorical modes of discourse in essay writing
Time: 2 minutes
Procedure:

1. Researcher then elicits different rhetorical modes of discourse in essay writing from participants and writes them on the board:
   Possible answers: *narrative/descriptive/expository/argumentative*
2. Researcher gives a definition or example of each mode.
3. Researcher points out that the aim of this course is how to write an argumentative essay.

Stage 4:
Aim: Overview of writing essay parts
Time: 5 minutes
Procedure:

1. Researcher quotes "*The paragraph [is] a mini-essay; it is also a maxi-sentence."* (Donald Hall).
2. Researcher shows the next slide containing a rectangle (four paragraph rectangle) and asks participants to imagine this is their sheet of paper for the essay. She then points to the four paragraphs signified by dotted lines and separated by indentation.
3. Researcher explains that a paragraph is a visual cue for readers. The indentation at the beginning, like the capital letter at the start of a sentence, signals the reader that a new thought unit is about to begin.
4. Researcher quotes *"The purpose of paragraphing is to give the reader a rest. The writer is saying...: Have you got that? If so, I'll go to the next point."* (H. W. Fowler)
5. Participants are asked to reflect on their writing experiences, and say what parts an essay should have:
   Possible answer: *Introduction/Body/Conclusion*
6. Researcher indicates that each of these paragraphs has a topic sentence which introduces the main idea of the paragraph.

7. Researcher explains that this four-session course will help participants to develop each paragraph of the essay effectively.

Stage 5:
Aim: Lead-in: Life without a plan!
Time: 5 minutes
Procedure:

1. Researcher asks participants what they would like her to do now.
2. Researcher listens to all their ideas and pretends to be confused and not able to decide.
3. Researcher then explains and elicits from them how it might feel not having a plan for a lesson.
4. Participants are led to the topic of planning as a crucial part of initiating writing.

Stage 6:
Aim: Introducing "**Planning" in writing**
Time: 5 minutes
Procedure:

1. Researcher begins by asking participants *why* they need to plan before starting to write and then shows some possible reasons on the next slide.
   Possible reasons:
   - *To put together the related and relevant ideas on the topic and develop an organized piece.*
   - *To remind you of important points that should be covered in your essay,*
   - *To highlight the final structure of your essay.*
2. Researcher then asks participants *what* they can plan before starting to write.
   Possible answers: *Time/Content/Language features*

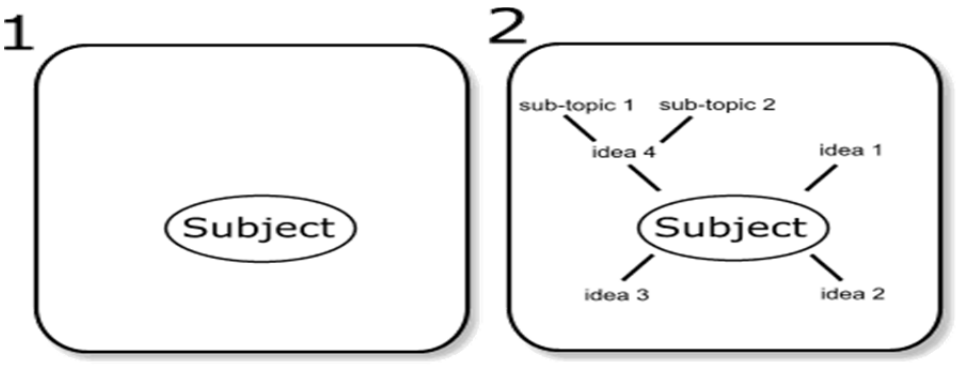
Stage 7:
Aim: Planning (Time)
Time: 5 minutes
Procedure:

1. Researcher explains that sometimes participants are required to have timed writing task. Thus managing time and allocating adequate time to each part of the essay is an important skill in writing an essay.
2. Researcher explains that depending on the whole time for the essay, the topic and length of the essay and the number of arguments to be included in the Body Paragraph, a writer should divide the time efficiently for each paragraph.

Stage 8:
Aim: Planning (Content)
Time: 5 minutes
Procedure:

1. Researcher explains that once participants read the essay topic, they need to analyse it carefully and give some thought to what their position (argument) could be.
2. Researcher explains that to initiate thinking and generate possible writing topics, it is important for participants to explore ideas for writing topics using a variety of pre-writing brainstorming strategies.

3. Researcher explains that brainstorming is simply the process of collecting ideas by thinking freely about all possibilities for a topic.
4. One of those possible brainstorming strategies is introduced: *Mind mapping*.
5. Researcher explains that mind mapping, or also called clustering, is a prewriting technique for those who are visually oriented because it allows them to generate and organize ideas in a visual context. It consists of a central word or concept around which a number of main ideas relating to that word are drawn by using circles and lines.
6. Researcher shows mind mapping through the following diagrammatical demonstration:



Stage 9:
Aim: Planning (Language Features)
Time: 15 minutes
Procedure:

1. Researcher explains one possible strategy to come up with a list of words and phrases that might be useful in the writing phase is brainstorming.
2. Researcher explains that when brainstorming, it is not necessary to keep writing continuously; they should just jot down ideas that seem related to their topic (like what they did in stage 9) and while doing so, they can come up with a list of words and phrases.
3. Researcher works on an example of generating ideas through brainstorming.
   Smoking should be banned from public. Yes? No?
   Some possible ideas for Yes are as follows:

   - *Smoking is dirty*
     - *leave ash*
     - *leave cigarette butts*
     - *smells bad*

   - *Smoking is dangerous*
     - *careless smokers cause fires*
     - *causes throat & lung cancer*
     - *causes heart disease*
     - *cause irritation to the eyes*

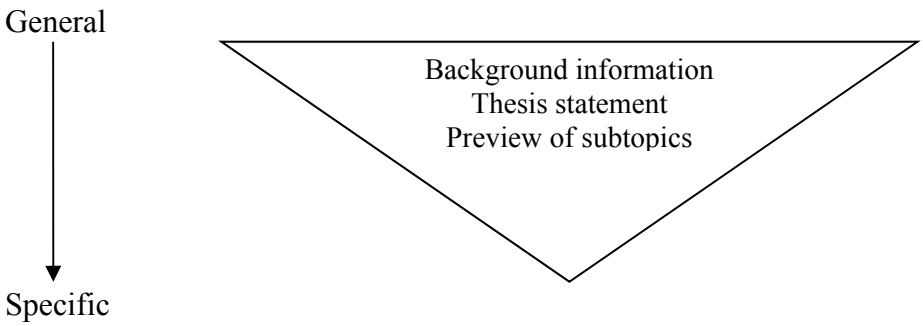
4. Researcher writes the following quotation on the board to emphasize more on the topic of outlining before writing:

"By writing an outline you really are writing in a way, because you're creating the structure of what you're going to do. Once I really know what I'm going to write, I don't find the actual writing takes all that long." (Tom Wolfe)

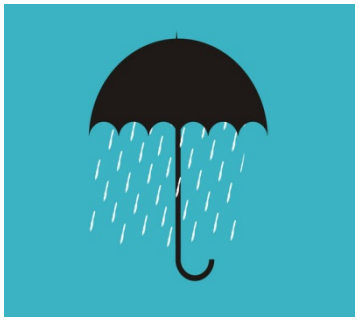
5. Participants are given another topic and told to work in groups and draw concept maps to share their ideas.

Stage 10:
Aim: Introduction Paragraph
Time: 15 minutes
Procedure:

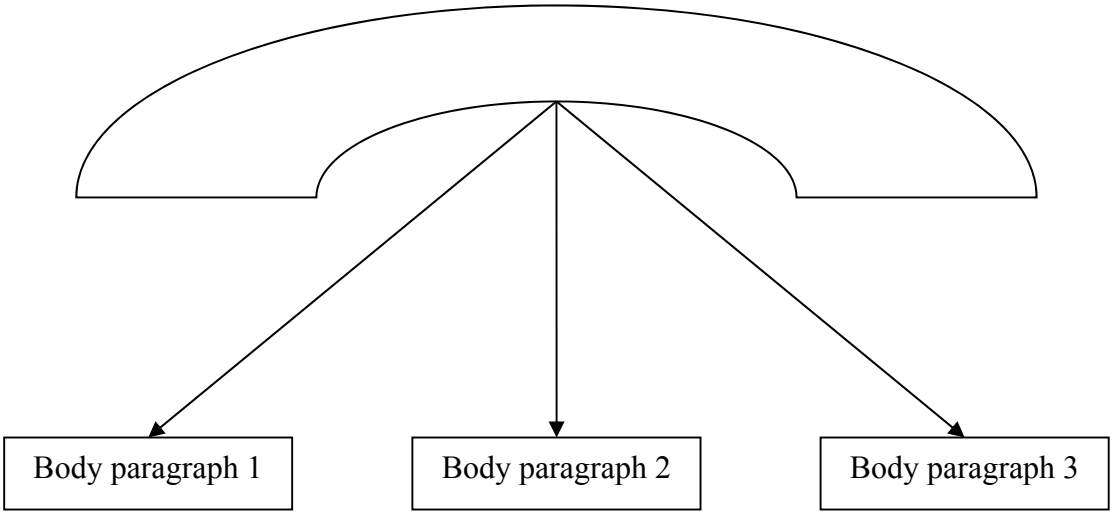
1. Researcher explains that after pre-writing planning, a writer needs to start writing the introduction as the first paragraph of the essay.
2. Researcher explains that there are many different ways a writer can begin an essay. Researcher writes some possible parts of the introduction paragraph on the board:
   - *Background information (a factual information, a recent news, a brief historical review, a summary of others' opinions)*
     [Attention grabber (usually a question or comment that inspires an emotional response from the reader. It should be used to get their interest.)]
   - *Thesis Statement (stating what the writer intends to argue)*
   - *Preview of subtopics*
3. Researcher explains that this structure is a general to specific format:



4. Researcher compares the introduction paragraph to an umbrella under which all other ideas fall.



5. Researcher states that what comes next is the body of the essay (the largest portion) in which the writer presents his/her points of discussion. Each point is typically given its own paragraph and is supported by evidence. Thus each paragraph contains only one key point (everything in the paragraph should support that one key point).



6. Researcher shows participants a sample introduction paragraph and asks them to label the three parts of the introductory paragraph in the model.
   A person born in the twentieth century has seen many changes take place in almost all areas of human life. Some people are excited by the challenges that these changes offer; others want to return to the simpler less automated lifestyle of the past. Living in the twentieth century certainly had many benefits but it also had many problems. This essay will discuss the higher standard of living which developed last century but also the increasingly polluted environment, the depersonalization of human relationships and the weakening of spiritual values.

7. Researcher shows participants another introduction paragraph and asks them to mark its different parts.
   People in different cultures all over the world have different systems for family life. In most cultures, people live in extended families, in which several generations share the same house. However, in others, the nuclear family is the norm, with only the parents and young children sharing the same house. In the United States, some people are experimenting with still another system of family life; living together without marriage. If this system becomes widespread, it could have enormous effects on American society. On the positive side, living together might reduce the divorce rate in the United States; on the negative side, it might lead to the eventual disintegration of the family altogether. The various arguments for and against are presented below.

Stage 11: Writing practice
Aim: Writing an essay
Time: 45 minutes
Procedure:

1. Researcher shows participants the last quotation of the day:
   "Practice, practice, practice writing. Writing is a craft that requires both talent and acquired skills. You learn by doing, by making mistakes and then seeing where you went wrong." (Jeffrey A. Carver)
2. Researcher then gives them a five-minute break time and then says it is time to practice writing and distributes the sheets of paper to participants and asks them to read the instruction and the topic carefully and write an argumentative essay.

APPENDIX C. Sample PowerPoint slides (Session 1)

IMPROVING EFFECTIVE ACADEMIC WRITING IN ENGLISH

Session 1 | Sara Amani

A Quotation

2

"Most people won't realize that writing is a craft. You have to take your apprenticeship in it like anything else."

(Katherine Anne Porter)

What is an Essay?

3

An essay is a **critical evaluation** which communicates an assemblage of **facts** and **opinions** about a certain **topic**.

Understanding the Title

4

Account for: *give the reasons for.*

Analyse: *examine in depth and describe the main characteristics of.*

Assess: *weigh up the elements of and arrive at a conclusion about.*

Comment: *give an opinion and provide evidence for your views.*

Compare: *bring out the similarities between.*

Contrast: *bring out the dissimilarities between.*

Define: *explain the exact meaning of.*

Describe: *use words and diagrams to illustrate.*

Discuss: *provide evidence or opinions about; arriving at a balanced conclusion.*

Understanding the Title

5

Evaluate: *weigh up or appraise.*

Explain: *make the meaning of something clear.*

Illustrate: *use diagrams or examples to make clear.*

Justify: *show that an idea or statement is correct.*

List: *provide an itemised series of statements about.*

Outline: *describe the essential parts only.*

Review: *examine critically.*

State: *express clearly.*

Summarise: *without illustrations, provide a brief account of.*

Rhetorical Modes of Discourse in Essay Writing

6

- ***Narrative**: telling a story,*
- ***Descriptive**: creating a vivid image of a person, place, or thing,*
- ***Expository**: telling how to make or do something, report on an experience, or explore an idea,*
- ***Argumentative**: stating opinions and supporting them convincingly.*

A Quotation

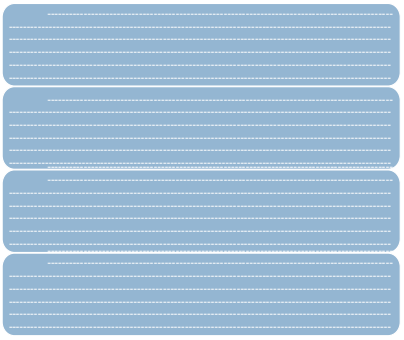
7

"The paragraph [is] a mini-essay; it is also a maxi-sentence."

(Donald Hall)

Visual Demonstration of Essay

8



A Quotation

9

"The purpose of paragraphing is to give the reader a rest. The writer is saying: Have you got that? If so, I'll go to the next point."

(H. W. Fowler)

Reflection

10

Reflect on your writing experiences, and say what parts an essay should have:

- *Introduction Paragraph*
- *Body Paragraph*
- *Conclusion Paragraph*

TREE

11

- ***Tell*** what you believe (e.g. Topic sentence),
- Provide two or more ***Reasons*** (e.g. why do I believe this?, Will my reader believe this?),
- ***End*** it, (i.e. wrap it up right),
- ***Examine*** (i.e. Do I have all my parts)

Life with/without a Plan!

12

"Planning" in Writing

13

Why do we need to plan before starting to write?

- *To put together the related and relevant ideas on the topic and develop an organized piece.*
- *To remind you of important points that should be covered in your essay,*
- *To highlight the final structure of your essay.*

"Planning" in Writing

14

What can we plan before starting to write?

- *Time*
- *Content*
- *Language features*

Planning (Time)

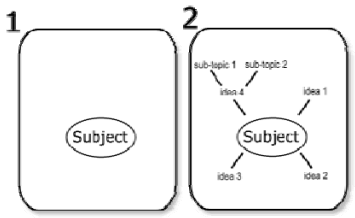
15

- *Timed writing task*
- *Managing time and* allocating adequate time to each part of the essay
- Depending on:

✓ whole time for the essay,
✓ topic of the essay,
✓ length of the essay,
✓ number of arguments in the Body Paragraph

Planning (Content)

16



Brainstorming(mind mapping):
A pre-writing strategy

Planning (Language Features)

17

Brainstorming (Listing): another pre-writing strategy

Smoking should be banned from public:

→ Smoking is dangerous
- careless smokers cause fires
- causes throat & lung cancer
- causes heart disease
- cause irritation to the eyes

→ Smoking is dirty
- leave ash
- leave cigarette butts
- smell bad

A Quotation

18

"By writing an outline you really are writing in a way, because you're creating the structure of what you're going to do. Once I really know what I'm going to write, I don't find the actual writing takes all that long."

(Tom Wolfe)

Introduction Paragraph

19

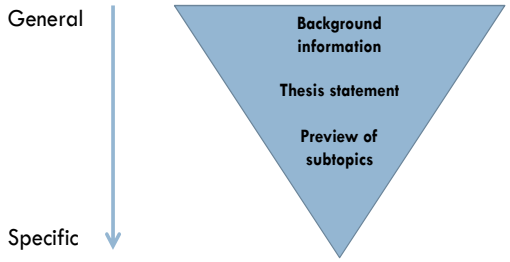
Structure of an Introduction Paragraph:

- ***Background Information*** *(a factual information, a recent news, a brief historical review, a summary of others' opinions)*

  *[**Attention Grabber** (usually a comment that inspires an emotional response from the reader. It should be used to get their interest.)]*
- ***Thesis Statement*** *(stating what the writer intends to argue)*
- ***Preview of subtopics***

General to Specific Structure

20

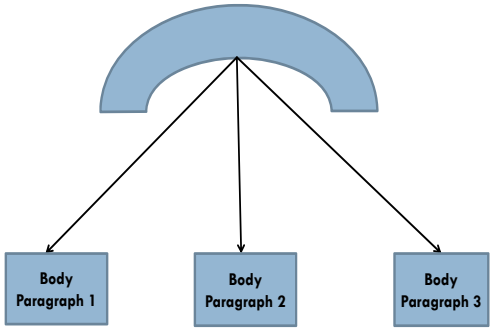


Introduction + Body

21

Body Paragraph

22



Sample Introduction Paragraph

23

A person born in the twentieth century has seen many changes take place in almost all areas of human life. Some people are excited by the challenges that these changes offer; others want to return to the simpler less automated lifestyle of the past. Living in the twentieth century certainly had many benefits but it also had many problems. This essay will discuss the higher standard of living which developed last century but also the increasingly polluted environment, the depersonalization of human relationships and the weakening of spiritual values.

Sample Introduction Paragraph

24

People in different cultures all over the world have different systems for family life. In most cultures, people live in extended families, in which several generations share the same house. However, in others, the nuclear family is the norm, with only the parents and young children sharing the same house. In the United States, some people are experimenting with still another system of family life; living together without marriage. If this system becomes widespread, it could have enormous effects on American society. On the positive side, living together might reduce the divorce rate in the United States; on the negative side, it might lead to the eventual disintegration of the family altogether. The various arguments for and against are presented below.

A Quotation

25

"Practice, practice, practice writing. Writing is a craft that requires both talent and acquired skills. You learn by doing, by making mistakes and then seeing where you went wrong."

(Jeffrey A. Carver)

What to do next?

26

Write an essay (45 minutes)

APPENDIX D. IELTS nine-band scale

Band 9 - Expert User Has fully operational command of the language: appropriate, accurate and fluent with complete understanding.
Band 8 - Very Good User Has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriacies. Misunderstandings may occur in unfamiliar situations. Handles complex detailed argumentation well.
Band 7 - Good User Has operational command of the language, though with occasional inaccuracies, inappropriacies and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.
Band 6 - Competent User Has generally effective command of the language despite some inaccuracies, inappropriacies and misunderstandings. Can use and understand fairly complex language, particularly in familiar situations.
Band 5 - Modest User Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field.
Band 4 - Limited User Basic competence is limited to familiar situations. Has frequent problems in understanding and expression. Is not able to use complex language.
Band 3 - Extremely Limited User Conveys and understands only general meaning in very familiar situations. Frequent breakdowns in communication occur.
Band 2 - Intermittent User No real communication is possible except for the most basic information using isolated words or short formulae in familiar situations and to meet immediate needs. Has great difficulty understanding spoken and written English.
Band 1 - Non User Essentially has no ability to use the language beyond possibly a few isolated words.
Band 0 - Did not attempt the test No assessable information provided

APPENDIX E. IELTS task 2 writing band descriptors

Band	Task Response	Cohesion and Coherence	Lexical Resource	Grammatical Range and Accuracy
9	- fully addresses all parts of the task - presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas	- uses cohesion in such a way that it attracts no attention - skilfully manages paragraphing	- uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	- uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	- sufficiently addresses all parts of the task - presents a well-developed response to the question with relevant, extended and supported ideas	- sequences information and ideas logically - manages all aspects of cohesion well - uses paragraphing sufficiently and appropriately	- uses a wide range of vocabulary fluently and flexibly to convey precise meanings - skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation produces rare errors in spelling and/or word formation	- uses a wide range of structures - the majority of sentences are error-free - makes only very occasional errors or inappropriacies
7	- addresses all parts of the task - presents a clear position throughout the response - presents, extends and supports main ideas, but there	- logically organises information and ideas; there is clear progression throughout - uses a range of cohesive devices appropriately	- uses a sufficient range of vocabulary to allow some flexibility and precision - uses less common lexical items with some awareness of style and	- uses a variety of complex structures - produces frequent error-free sentences - has good control of grammar and punctuation but may make a few errors

	may be a tendency to overgeneralise and/or supporting ideas may lack focus	although there may be some under/over-use - presents a clear central topic within each paragraph	collocation - may produce occasional errors in word choice, spelling and/or word formation	
6	- addresses all parts of the task although some parts may be more fully covered than others - presents a relevant position although the conclusions may become unclear or repetitive - presents relevant main ideas but some may be inadequately developed/ unclear	- arranges information and ideas coherently and there is a clear overall progression - uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical - may not always use referencing clearly or Appropriately - uses paragraphing but not always logically	-uses an adequate range of vocabulary for the task - attempts to use less common vocabulary but with some inaccuracy - makes some errors in spelling and/or word formation, but they do not impede communication	- uses a mix of simple and complex sentence forms - makes some errors in grammar and punctuation but they rarely reduce communication
5	- addresses the task only partially; the format may be inappropriate in places - expresses a position but the development is not always clear and there may be no conclusions drawn - presents some	- presents information with some organisation but there may be a lack of overall progression - makes inadequate, inaccurate or over-use of cohesive devices - may be repetitive because	- uses a limited range of vocabulary, but this is minimally adequate for the task - may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	- uses only a limited range of structures - attempts complex sentences but these tend to be less accurate than simple sentences - may make frequent grammatical errors and punctuation may be faulty;

	main ideas but these are limited and not sufficiently developed; there may be irrelevant detail	of lack of referencing and substitution -may not write in paragraphs, or paragraphing may be inadequate		errors can cause some difficulty for the reader
4	- responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate - presents a position but this is unclear - presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported	- presents information and ideas but these are not arranged coherently and there is no clear progression in the response - uses some basic cohesive devices but these may be inaccurate or repetitive - may not write in paragraphs or their use may be confusing	- uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task - has limited control of word formation and/or spelling; errors may cause strain for the reader	- uses only a very limited range of structures with only rare use of subordinate clauses - some structures are accurate but errors predominate, and punctuation is often faulty
3	- does not adequately address any part of the task - does not express a clear position - presents few ideas, which are largely undeveloped or irrelevant	- does not organise ideas logically - may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	- uses only a very limited range of words and expressions with very limited control of word formation and/or spelling - errors may severely distort the message	- attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	- barely responds to the task - does not express a position - may attempt to present one or two	- has very little control of organisational features	- uses an extremely limited range of vocabulary; essentially no control of word formation and/or	- cannot use sentence forms except in memorised phrases

	ideas but there is no development		spelling	
1	- answer is completely unrelated to the task	- fails to communicate any message	- can only use a few isolated words	- cannot use sentence forms at all
0	- does not attend - does not attempt the task in any way - writes a totally memorised response			

APPENDIX F. Extra tables of results in Study 1 & 2

Table 1 ***Descriptive statistics for recruitment test in Study 1***

	N	Mean	Std. Deviation	Std. Error
+M+P	10	5.15	.24	.07
+M-P	10	5.30	.42	.13
-M+P	8	5.18	.53	.18
-M-P	7	5.21	.48	.18
Total	35	5.21	.40	.06

Note. M: Metacognitive instruction, P: Planning

Table 2 ***ANOVA: Error-free clauses in Study 1***

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	.01	3	.00	.35	.78
	Within Groups	.48	31	.01		
	Total	.49	34			
Post-test	Between Groups	.01	3	.00	.39	.75
	Within Groups	.34	31	.01		
	Total	.35	34			

*. The mean difference is significant at the .05 level.

Table 3 ***ANOVA: Gain scores for error-free clauses in Study 1***

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.06	3	.02	.62	.60
Within Groups	1.13	31	.03		
Total	1.20	34			

Table 4 ***Paired-sample t-test: Error-free verb forms in Study 1***

		Paired Differences					
		Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
+M+P	Pre-test - Post-test	-.01	.08	.02	-.49	9	.63
+M-P	Pre-test - Post-test	-.01	.10	.03	-.47	9	.64
-M+P	Pre-test - Post-test	-.03	.10	.03	-.88	7	.40
-M-P	Pre-test - Post-test	.12	.18	.06	1.84	6	.11

Note. M: Metacognitive instruction, P: Planning

*. The mean difference is significant at the .05 level.

Table 5 ***ANOVA: Syntactic complexity in Study 1***

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	.52	3	.17	1.04	.38
	Within Groups	5.12	31	.16		
	Total	5.64	34			
Post-test	Between Groups	.20	3	.06	.36	.78
	Within Groups	5.75	31	.18		
	Total	5.95	34			

*. The mean difference is significant at the .05 level.

Table 6 ***ANOVA: Lexical variety in Study 1***

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	3.55	3	1.18	2.33	.09
	Within Groups	15.75	31	.50		
	Total	19.30	34			
Post-test	Between Groups	1.15	3	.38	.67	.57
	Within Groups	17.73	31	.57		
	Total	18.88	34			

*. The mean difference is significant at the .05 level.

Table 7 ***Descriptive statistics for recruitment test in Study 2***

		N	Mean	Std. Deviation	Std. Error
Rater 1	+M+P	14	4.85	.30	.08
	+M-P	13	4.88	.41	.11
	-M+P	14	4.85	.41	.11
	-M-P	14	5.03	.13	.03
	Control Group	15	4.70	.25	.06
	Total	70	4.86	.32	.04
Rater 2	+M+P	14	4.75	.26	.07
	+M-P	13	4.92	.45	.12
	-M+P	14	4.93	.38	.10
	-M-P	14	5.00	.19	.05
	Control Group	15	4.76	.32	.08
	Total	70	4.87	.33	.04

Note. M: Metacognitive instruction, P: Planning

Table 8 ***ANOVA: Error-free clauses in Study 2***

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	.03	4	.00	.70	.60
	Within Groups	.83	65	.01		
	Total	.87	69			
Immediate Post-test	Between Groups	.05	4	.01	.78	.54
	Within Groups	1.06	65	.01		
	Total	1.11	69			
Delayed Post-test	Between Groups	.05	4	.01	1.30	.28
	Within Groups	.68	65	.01		
	Total	.74	69			

*. The mean difference is significant at the .05 level.

Table 9 ***ANOVA: Error-free verb forms in Study 2***

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	.07	4	.02	1.31	.27
	Within Groups	.94	65	.01		
	Total	1.01	69			
Immediate Post-test	Between Groups	.05	4	.01	.78	.54
	Within Groups	1.06	65	.01		
	Total	1.11	69			
Delayed Post-test	Between Groups	.07	4	.02	1.56	.19
	Within Groups	.76	65	.01		
	Total	.83	69			

*. The mean difference is significant at the .05 level.

Table 10 ***ANOVA: Syntactic complexity in Study 2***

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	.10	4	.02	.17	.95
	Within Groups	9.54	65	.14		
	Total	9.64	69			
Immediate Post-test	Between Groups	.21	4	.05	.31	.87
	Within Groups	11.16	65	.17		
	Total	11.38	69			
Delayed Post-test	Between Groups	1.25	4	.31	1.44	.23
	Within Groups	14.18	65	.21		
	Total	15.44	69			

*. The mean difference is significant at the .05 level.

Table 11 ***ANOVA: Syntactic variety in Study 2***

		Sum of Squares	df	Mean Square	F	Sig.
Pre-test	Between Groups	13.83	4	3.46	1.64	.17
	Within Groups	136.75	65	2.10		
	Total	150.58	69			
Immediate Post-test	Between Groups	15.43	4	3.86	1.98	.10
	Within Groups	126.40	65	1.94		
	Total	141.84	69			
Delayed Post-test	Between Groups	15.68	4	3.92	1.50	.21
	Within Groups	170.08	65	2.61		
	Total	185.771	69			

*. The mean difference is significant at the .05 level.

APPENDIX G. Ethics forms

Department of Applied Language Studies and Linguistics
Fisher Building, 18 Waterloo Quadrant
Telephone: +64-9-373-7599

The University of Auckland
Private Bag 92019
Auckland, New Zealand

PARTICIPANT INFORMATION SHEET

To: Participants

Title: Improving effective academic writing in English

My name is Sara Amani. I am enrolled as a PhD student in the Department of Applied Language Studies and Linguistics at the University of Auckland. I am conducting this research for my doctoral thesis in which I am investigating the ways in which second language learners of English may benefit from instruction and practice in argumentative writing tasks. I would like to invite you to participate in my research and I would highly appreciate your cooperation.

What does it involve?

My research is a classroom-based study which involves questionnaires, use of writing practice and tests. Audio-recordings of think-aloud activities will also be collected from 5 participants on a voluntary basis. Over a period of 6 weeks, you will complete 11 hours on effective writing. This includes:

A. IELTS writing task to establish your suitability for participation (45 min) + Writing strategy questionnaire (15 min)

B. After recruitment:
4 sessions (2 hours each): Writing instruction (1 hour) + argumentative writing practice (45 min) + [think-aloud activity from only 5 participants] + reflective questionnaire (15 min)

C. An argumentative writing task (45 min) + a writing strategy questionnaire (15 min)

D. 3 weeks later: An argumentative writing task (45 min) + a writing strategy questionnaire (15 min)

How will it benefit you?

I hope that this study will have benefits for your language learning, both in providing opportunities to receive writing instruction, and practice in argumentative writing in your second language, which will be beneficial to your university studies. I will provide feedback on individual task performance at the end of the study to those who would like to receive a summary of results.

What happens to the information collected?

The writing tasks will be scored, the questionnaires will be coded and think-aloud recordings will be transcribed by me only. You will be offered the opportunity to edit the transcripts of your recordings when the data have been transcribed. You may also ask for an electronic copy of your recordings. Codes will replace your names, thus it should be explained that your name will not be identifiable when I write up my research reports for the doctoral thesis, journal publications, or presentations at conferences. All the data will be stored in a locked cabinet for a period of six years after the completion of the study and used for possible further development of the current research. After this time, all information will be shredded by following the usual procedure of destroying confidential data at the university.

Your choice

Participation in this research is entirely voluntary. Your Head of Department has given assurance that your participation or non-participation in this study will not affect your relationship with the school or course grades and you are free to withdraw from the project at any time or withdraw your data at any time up to 1 October 2010, without giving reasons for your withdrawal.

If you agree to participate, please kindly sign the Consent Form and return it to me. If you have any queries or wish to have more information, please contact me:

Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland
New Zealand
Email: sama014@aucklanduni.ac.nz
NZ Mobile: 0064 211790350
Australia Phone: To be confirmed

My supervisor is:
Dr. Jenefer Philp

Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland
Email: j.philp@auckland.ac.nz
Phone: 3737599 Ext. 84967

The Head of Department is:
Professor Yan Huang
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland
Email: yan.huang@auckland.ac.nz
Phone: 3737599 Ext. 87809

For any queries regarding ethical concerns you may contact:
The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone: 09 373-7599 extn. 83711.

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 26 JULY 2010 for (3) years, Reference Number 2010/318

Department of Applied Language Studies and Linguistics
Fisher Building, 18 Waterloo Quadrant
Telephone: +64-9-373-7599

The University of Auckland
Private Bag 92019
Auckland, New Zealand

CONSENT FORM FOR PARTICIPANTS

THIS CONSENT FORM WILL BE HELD FOR A PERIOD OF SIX YEARS

Title: Improving effective academic writing in English
Researcher: Sara Amani

- I have read the Participant Information Sheet, have understood the nature of the research and why I have been selected. I have had the opportunity to ask questions and have them answered to my satisfaction.
- I understand that the Head of Department has given permission for this study.
- I understand that the Head of Department has given assurance that my participation or non-participation in this study will not affect my relationship with the school or my course grades.
- I understand that I may be precluded from participation on the basis of my performance on the initial IELTS writing task.
- I understand that I am free to withdraw from the project at any time or withdraw my data at any time up to 1 October 2010, without giving reasons for my withdrawal.
- I understand that neither my name nor the university name will be identifiable in the research report.
- I understand that data will be stored for 6 years in a locked cabinet on university premises and will then be destroyed by shredding the hard copies.
- I agree / do not agree to be audio-taped.
- I wish / do not wish to edit the transcripts of my recordings.
- I wish / do not wish to receive an electronic copy of my recordings.
- I wish / do not wish to receive the summary of findings.
-

Name _____

Signature _____ Date _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 26 JULY 2010 FOR (3) YEARS REFERENCE NUMBER 2010/318

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

Department of Applied Language Studies and Linguistics
Fisher Building, 18 Waterloo Quadrant
Telephone: +64-9-373-7599

The University of Auckland
Private Bag 92019
Auckland, New Zealand

PARTICIPANT INFORMATION SHEET

To: Program Head/Course Coordinator: English as an International Language

Title: Improving effective academic writing in English

My name is Sara Amani. I am enrolled as a PhD student in the Department of Applied Language Studies and Linguistics at the University of Auckland. I am conducting this research for my doctoral thesis in which I am investigating the possibility of enhancing L2 learners' written performance on argumentative tasks by raising their metacognitive awareness and providing them with adequate pre-task planning. I would like to seek your permission to invite students, on a voluntary basis, to participate in my research, and I would appreciate any assistance you can offer to me.

What does it involve?

My research is a classroom-based study of 140 participants: 60 for the initial pilot study and 80 in the main study. In both studies, the participants are divided into 4 equal groups. It involves questionnaires, use of writing practice and tests. Audio-recordings of think-aloud activities will also be collected from 5 participants on a voluntary basis. Dates and time for data collection will be set in accordance with the schedules of the targeted students. Over a period of 6 weeks, the students will complete 11 hours on effective writing. This includes:

A. IELTS writing task to establish the suitability of students for participation (45 min) + Writing strategy questionnaire (15 min)

B. After recruitment:
4 sessions (2 hours each): Writing instruction (1 hour) + argumentative writing practice (45 min) + [think-aloud activity from only 5 participants] + reflective questionnaire (15 min)

C. An argumentative writing task (45 min) + a writing strategy questionnaire (15 min)

D. 3 weeks later: An argumentative writing task (45 min) + a writing strategy questionnaire (15 min)

How will it benefit participants?

The benefits to participants include opportunities to receive writing instruction and practice in argumentative writing in their second language, which will be beneficial to their university studies. I will provide feedback on individual task performance at the end of the study to those who would like to receive a summary of results.

What happens to the information collected?

The writing tasks will be scored, the questionnaires will be coded and think-aloud recordings will be transcribed by me only. Participants will be offered the opportunity to edit the transcripts of their recordings when the data have been transcribed. They may also ask for an electronic copy of their recordings. Codes will replace participants' names, thus it should be explained that their names will not be identifiable when I write up my research reports for the doctoral thesis, journal publications, or presentations at conferences. All the data will be stored in a locked cabinet for a period of six years after the completion of the study and used for possible further development of the current research. After this time, all information will be shredded by following the usual procedure of destroying confidential data at the university.

It should also be explained that the participants are free to withdraw from the project at any time or withdraw their data at any time up to 1 October 2010, without giving reasons for withdrawal. You may also withdraw your permission for the recruitment of the project at any time or withdraw any data traceable to the study in the department at any time up to 1 October 2010, without giving reasons.

Thank you for your time to consider this request. I would be grateful if you could give me permission to conduct the research within your program. This research will only be conducted within your program but your university will not be named in my research reports or publications. I would ask for your assurance that participation or non-participation of the students will not affect their relationship with the school and their course grades in any way.

If you agree to the participation of students, please kindly sign the Consent Form and return it to me. If you have any queries or wish to have more information please contact me:

Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland
New Zealand
Email: sama014@aucklanduni.ac.nz
NZ Mobile: 0064 211790350
Australia Phone: To be confirmed

My supervisor is:
Dr. Jenefer Philp
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland
Email: j.philp@auckland.ac.nz
Phone: 3737599 Ext. 84967

The Head of Department is:
Professor Yan Huang
Department of Applied Language Studies and Linguistics
The University of Auckland
Private Bag 92019
Auckland
Email: yan.huang@auckland.ac.nz
Phone: 3737599 Ext. 87809

For any queries regarding ethical concerns you may contact:
The Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone: 09 373-7599 extn. 83711.

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 26 JULY 2010 for (3) years, Reference Number 2010/318

Department of Applied Language Studies and Linguistics
Fisher Building, 18 Waterloo Quadrant
Telephone: +64-9-373-7599

The University of Auckland
Private Bag 92019
Auckland, New Zealand

CONSENT FORM FOR PROGRAM HEAD/COURSE COORDINATOR
THIS CONSENT FORM WILL BE HELD FOR A PERIOD OF SIX YEARS

Title: Improving effective academic writing in English

Researcher: Sara Amani

- I have read the Participant Information Sheet, have understood the nature of the research and have given permission for this study.
- I understand that I am free to withdraw permission for the recruitment of the project at any time or withdraw any data traceable to the study in the department at any time up to 1 October 2010, without giving reasons for the withdrawal.
- I give my assertion that participation/non participation of students will not affect their relationship with the school or their course grades.
- I understand that neither the participants' names nor the university name will be identifiable in the research report.
- I understand that data will be stored for 6 years in a locked cabinet on university premises and will then be destroyed by shredding the hard copies.

Name _____

Signature _____ Date _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 26 JULY 2010 FOR (3) YEARS REFERENCE NUMBER 2010/318