Guided planning, task complexity and second language oral development

by

Colin Thompson

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire

March 2014

Abstract

This thesis reports on a mixed methods experimental research study carried out at a university in Japan. The study investigated the effectiveness of two types of guided planning treatment towards specific language forms. Specifically, English relative clause types OS and OPREP as well as 3rd person singular and plural. Two groups of Japanese second year intermediate level learners performed a series of oral narrative tasks that increased in complexity over a three week period. Both groups were placed under different planning conditions. One condition involved 'guided planning' which consisted of continuous guidance towards English relative clauses and 3rd person singular and plural. The other condition 'guided and unguided planning' consisted of initial guidance towards the target forms and then the learners received unguided planning during the rest of the task sequence. During the treatment, both groups were interviewed about their planning strategies.

It was hypothesized that the guided planning group would produce greater developmental gains in accuracy compared to the guided and unguided planning group. Learners' L2 speech was measured in terms of fluency, accuracy and complexity. The results showed that the guided planning group produced significantly greater gains in fluency and accuracy compared to the guided and unguided planning group. In addition, both groups focused on form during the task sequencing treatment.

No previous studies have appeared to investigate the effects of guided and unguided planning with tasks that are sequenced over time. As a result, the findings of this study appear unique in reporting the benefits that guided planning and task complexity produces on L2 oral development in terms of fluency, accuracy and complexity.

Table of Contents

List of tables and illustrations

List of Figures

Acknowledgements

I'd like to thank everyone who has helped me with this thesis. Firstly, a big thank you to Christian Jones for being an excellent supervisor and for being really supportive and helpful throughout this study. Thank you to all the staff at UCLAN. I've really enjoyed my time as a student and the feedback and friendly support I've always received has made my study experience memorable. In addition, I'd like to thank colleagues outside of UCLAN who have assisted me in this thesis, Dr Peter Robinson for all his advice and guidance on task complexity and TBLT. Thank you to Mayya Levkina for all her helpful advice with data analysis and the use of CLAN. Finally, I'd like to thank my family and close friends for all their support during this thesis. Especially my wife Yuki for all her super support whilst I've been studying!

This thesis is dedicated to my son, Luke. I hope we have a lot of fun together after finishing this study!

Abbreviations

APU	Ritsumeikan Asia Pacific University.
B1GP	A learner of B1 oral proficiency according to the CEFR (see below) who receives guided planning (GP) before performing an oral task.
B2GP	A learner of B2 oral proficiency according to the CEFR who receives guided planning before performing an oral task.
B1UP	A learner of B1 oral proficiency according to the CEFR (see below) who receives unguided planning (UP) before performing an oral task.
B2UP	A learner of B2 oral proficiency according to the CEFR who receives unguided planning before performing an oral task.
CAF	Complexity, Accuracy, Fluency. This term is used to represent three aspects of L2 oral production.
CAQDAS	Computer assisted qualitative data analysis.
CEFR	The Common European Framework of References for Languages.
CLT	Communicative Language Teaching.
CTP	Communicational Teaching Project.
EFL	English as a foreign language. This term refers to learners who are studying a second language (English) in their native country.
GP	Guided planning. This term is to refer to learners who receive assistance towards grammar before performing an oral task.
GUP	Guided and unguided planning. This term refers to learners who receive a combination of grammar guidance as well as unguided planning time before performing an oral task.
JET	The Japan Exchange and Teaching Programme.
KR-20	Kuder Richardson 20. This term refers to computer software used to measure the reliability of test items.
L1	A person's native language.
L2	A person's second language.
MEXT	The Japanese Ministry of Education, Culture, Sports, Science and Technology.
NP	Control Group. This term refers to learners who do not take part in any of the task planning and sequencing treatment of the pilot study.

OO	Object object. This term is considered to be a simple relative clause type.
OS	Object subject. This term is considered to be a simple relative clause type.
OPREP	Object of a preposition. This term is considered to be a difficult relative clause type.
RCs	English relative clauses.
SLA	Second language acquisition.
TBLT	Task-based language teaching.
TOEFL	Test of English as a foreign language.
TSLT	Task-supported language teaching.
UP	Unguided planning. This term refers to learners who receive unguided planning time before performing an oral task in the pilot study.

1. INTRODUCTION

1.1 Introduction

This chapter is intended to provide a brief overview of this thesis in terms of background, aims, methodology and findings. The past twenty years has since a significant amount of research on the role of task planning as a means for developing learners' second language (L2) speaking skills. One type of task planning known as strategic planning takes place before the performance of a task, when learners are provided with instructions and are given time to prepare (Ellis, 2005). Task planning research has shown generally consistent results in relation to strategic planning and its impact on L2 performance (for example, Gilbert, 2007b; Mochizuki & Ortega, 2008; Skehan & Foster, 2005). Numerous studies such as Yuan & Ellis (2003); Kawauchi, (2005) have confirmed that when learners engage in strategic planning, they can speak in the L2 with greater fluency and greater complexity whilst Mochizuki & Ortega (2008) showed that strategic planning which included grammar assistance, referred to as guided planning led to improvements in learners' accuracy. A limitation of these studies however, is that they only address the immediate effects of task planning i.e. they involve 'one-off' experiments that examine task planning at a specific point in time (Ellis, 2005, 2009a). Consequently, there appears to be no strategic planning studies that have investigated L2 oral development of linguistic forms over time (Ellis, 2009a). In addition to the benefits of strategic planning, only a few studies have investigated the strategies learners use during planning (for example, Sangarun, 2005; Ortega, 2005; Kawauchi, 2005). Ortega's (2005) study showed that cultural and social factors influence how Spanish learners of English plan for oral tasks which affected their oral performance in terms of fluency and accuracy. However, Samuda & Bygate (2008) note that no studies have appeared to report the strategies learners use as they plan for tasks *over time*. Such findings would be pedagogically useful in knowing how learners attend to different aspects of their speech on subsequent task performances.

1.2 Research Objectives

This study was carried out to the fill the gaps concerning the lack of longitudinal task planning research outlined above. In addition, it attempts to show how tasks can be sequenced to develop Japanese learners' L2 oral skills within a Japanese educational context. Since 2003, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan has been concerned with traditional methods of English language instruction that have focused heavily on grammar translation, reading and writing which have not been seen as effective for improving Japanese university learners' L2 oral skills. Japanese students generally receive six years of English language instruction at junior and high school level prior to entering university. However, these lessons generally focus on reading, writing and grammar at the expense of fostering oral communication skills (Browne & Kichuchi, 2009). One reason for a reliance on traditional methods of language teaching in Japan concerns university entrance exams which do not evaluate L2 speaking and are instead grammar-focused. Consequently, "to get the important job done of preparing their students for university entrance exams, which mainly test English reading skills and knowledge of grammar and vocabulary, many Japanese teachers choose to teach grammar at the expense of communication" (Mochizuki & Ortega, 2008, p. 12). Studies that have investigated the perceptions of Japanese teachers of English regarding communicative language teaching (CLT) have shown that CLT is not considered beneficial for grammar learning therefore teachers have been reluctant to embrace it (O'Donnell, 2005; Sakui, 2004).

In terms of Japanese learners' exposure to native English speaking teachers, the Japan Exchange and Teaching Programme (JET) was initiated by the Japanese Government with the aim of recruiting native English teachers to work in Japanese secondary schools. According to Sakui (2007) "exposure to these native speaker teachers is, however, minimal and the system is not free from criticism" (p. 44). For example, one issue relates to the lack of teacher training as applicants can apply for the programme without having a teaching qualification. Thus, given Japanese learners' apparent lack of exposure with native English teachers, as well as issues relating to large class sizes that can contain 30 to 40 students, Mochizuki & Ortega, 2008 claim:

> learners can be expected to encounter very few opportunities for individually tailored communicative and form-focused experiences, if the responsibility for

> initiating and monitoring such one-on-one experiences is solely placed on the already overburdened classroom teachers or on the rarely available target language (L2) users outside the classroom. (p. 12)

Furthermore, Sakui (2004, 2007) notes that large class sizes hinder attempts to develop learners' L2 oral skills due to classroom management difficulties as teachers are often unable to effectively monitor multiple pairs or groups of students interacting in the L2 which can result in students reverting back to their L1 during communication. Teachers have therefore preferred to focus on other skills such as listening and grammar exercises which are considered easier to manage. Consequently, the above issues have resulted in many Japanese learners entering university having had little practice using their spoken English in communicative situations during their education (Browne & Kichuchi, 2009).

Efforts have been made at the university level to improve Japanese learners' L2 speech through the use of oral tasks (for example, Robinson, 2001, 2007; Thompson & Millington, 2012). However, task-based language teaching (TBLT) has been problematic within a Japanese educational context. For example, advocates of TBLT such as Willis (1996) favour using oral tasks to engage learners in meaningful language use upon which teachers can focus on developing learners' language after task performance. In other words, the instruction of language occurs *after* learners complete tasks. However, this method of language teaching has proved difficult to implement within Japanese university classrooms. Ellis (2009b) notes that most educational institutions rely on structural approaches to language learning in which grammatical features are first instructed and then practiced with exercises or activities. A structural approach towards language learning is not compatible with TBLT in which attention to form occurs after communication. As a result, the majority of TBLT studies to date have been conducted in experimental settings outside of university course programs (Robinson, 2011). Furthermore, given Japanese learners' previous educational background which lacks attention to L2 speaking, students may feel reluctant to engage in oral communication tasks in which they are required to interact in L2 without receiving any language guidance prior to their performance.

Given Japanese learners exposure to more traditional grammar translation methods of instruction that lack an emphasis on speaking, the position I take in this thesis seeks to argue the case for using oral tasks that provide instruction to language form *prior* to performance. Specifically, this involves the use of guided planning, otherwise known as task-supported language teaching (TSLT), which may provide

conditions that could successfully facilitate the development of Japanese learners' oral skills. In addition, guided planning also has the added advantage of drawing learners' attention towards linguistic forms known for their difficulty in oral production and encouraging its use during task performance. In the case of Japanese learners, a linguistic feature known for its difficulty in oral production is English relative clauses. Mochizuki & Ortega (2008) appears to be the only study that has targeted relative clauses in natural language use through guided planning, however no task planning studies have appeared to investigate learners' development of the form over time. In order to do so, a theoretically grounded proposal for sequencing tasks is needed to maximise L2 oral development. One possibility is the use of task complexity which involves sequencing tasks according to an increase in their cognitive demands. Robinson (2010) argues that sequencing tasks from simple to complex serves to push learners' output and provides optimal conditions for promoting L2 development in terms of fluency, accuracy and complexity. However, no studies have appeared to combine the effects of guided planning with tasks that increase in cognitive complexity over time. Furthermore, as outlined earlier, no research has appeared to investigate the strategies learners use when planning for tasks over time such as preparing for tasks that increase in complexity. Consequently, the aim of this study is to investigate the following gaps in previous research:

1. Combine the effects of guided planning and task complexity in order to provide task sequencing treatment that could maximise L2 oral development in terms of fluency, accuracy and complexity over time.
2. Develop Japanese learners' accuracy of specific types of English relative clauses (OS and OPREP) as well as the linguistic features that can accompany them such as 3rd person singular or plural for example, '*he likes the dog which has long hair'*.
3. Investigate the strategies Japanese learners use when planning for oral tasks that increase in complexity in order to provide an insight into the cognitive processes learners engage in as they prepare for tasks over time.

Given these reasons, this study attempts to answer the following two research questions:

- To what extent does guided planning and task complexity facilitate L2 oral development in terms of fluency, morphological accuracy involving OS and OPREP English relative clauses and 3rd person singular and plural, and syntactic complexity of second year Japanese university learners of English?
- What strategies do Japanese second year university learners of English use when planning for oral narratives that increase in complexity over time?

1.3 Methodology outline and findings

This study involved a pre- post-test design that lasted a total of seven weeks in which two groups of learners performed a sequence of oral narratives that increased in complexity over a three week period under different planning conditions. One group of learners received guidance towards relative clauses and 3rd person singular and plural as they prepared for tasks that increased in complexity, referred to as guided planning and task complexity (GP). The other group of learners received initial guidance towards the targeted forms during week one but were then free to plan independently on subsequent tasks that increased in complexity during weeks two and three, referred to as guided and unguided planning and task complexity (GUP). During the task sequencing treatment, qualitative research was carried out in the form of post-task interviews in order to investigate the strategies learners used as they prepared for more complex tasks over time.

The unique aspect about the findings of this study is that it appears to be the only study that has investigated the effects of guided planning over time. The results showed that guided planning and task complexity, as well as guided and unguided planning and task complexity, produced significant gains in fluency, accuracy and complexity. In addition, guided planning and task complexity produced significantly greater gains in terms of fluency and accuracy compared to guided and unguided planning and task complexity. The findings demonstrated that guided planning which involved explicit instruction towards OS and OPREP relative clauses as well as 3rd person singular and plural resulted in explicit learning of the forms from both groups whilst practice opportunities using the forms with more complex tasks resulted in the proceduralisation of the target language. The significant developments of both groups' L2 speech as a result of their respective task sequencing conditions, particularly guided

planning and task complexity points to the pedagogic benefits of sequencing tasks that combine guided planning and task complexity to improve fluency, accuracy and complexity. In addition, the findings of this study suggest the potential contribution that task-supported language teaching can provide within a Japanese educational context as a suitable means for improving Japanese learners' L2 oral skills.

In terms of research question two, the findings show that there were a lot of similarities between both groups' strategies as they planned for oral tasks over time even though they operated under different planning conditions. At the start of the treatment when both groups were provided with explicit guided instruction towards the targeted grammar forms, both groups focused on form during strategic planning. This planning strategy remained largely unchanged for the GP group as they continued to receive guided planning throughout the task sequencing treatment. The GUP group however, received unguided planning during weeks two and three of the task sequencing treatment yet they still largely maintained a focus on the targeted grammar points during this period whilst also showing evidence of attention towards form-in-meaning as certain learners focused on the storyline as well. In summary, both groups appeared to acknowledge the value of the grammar guidance provided in helping them meet the demands of narrative tasks that required its use, and therefore consciously attended towards practicing the forms throughout their respective task sequencing treatments.

1.4 Conclusion

The purpose of this chapter was to provide a brief outline of the present study in terms of background, research aims, methodology, and the research contribution of the findings. It is hoped that this study will encourage future research into the effects of guided planning and task complexity as a means for promoting L2 oral development as well as providing a guideline for teachers on how tasks can be designed and sequenced to facilitate L2 oral development within in a Japanese educational context. This thesis begins with chapter two which provides a backdrop into the research concerning L1 and L2 speech production before moving on to discuss L2 fluency, accuracy and complexity. Chapter three then describes the literature surrounding task planning and task complexity. Chapter four reports on a pilot study whilst chapter five describes and

justifies the methodology of the present study. Chapter six analyses the results relating to research question one whilst chapter seven analyses the results concerning research question two. Chapter eight discusses the findings of the thesis and finally, chapter nine reports the conclusions and limitations of the study as well as areas for future research.

2 LITERATURE REVIEW PART ONE

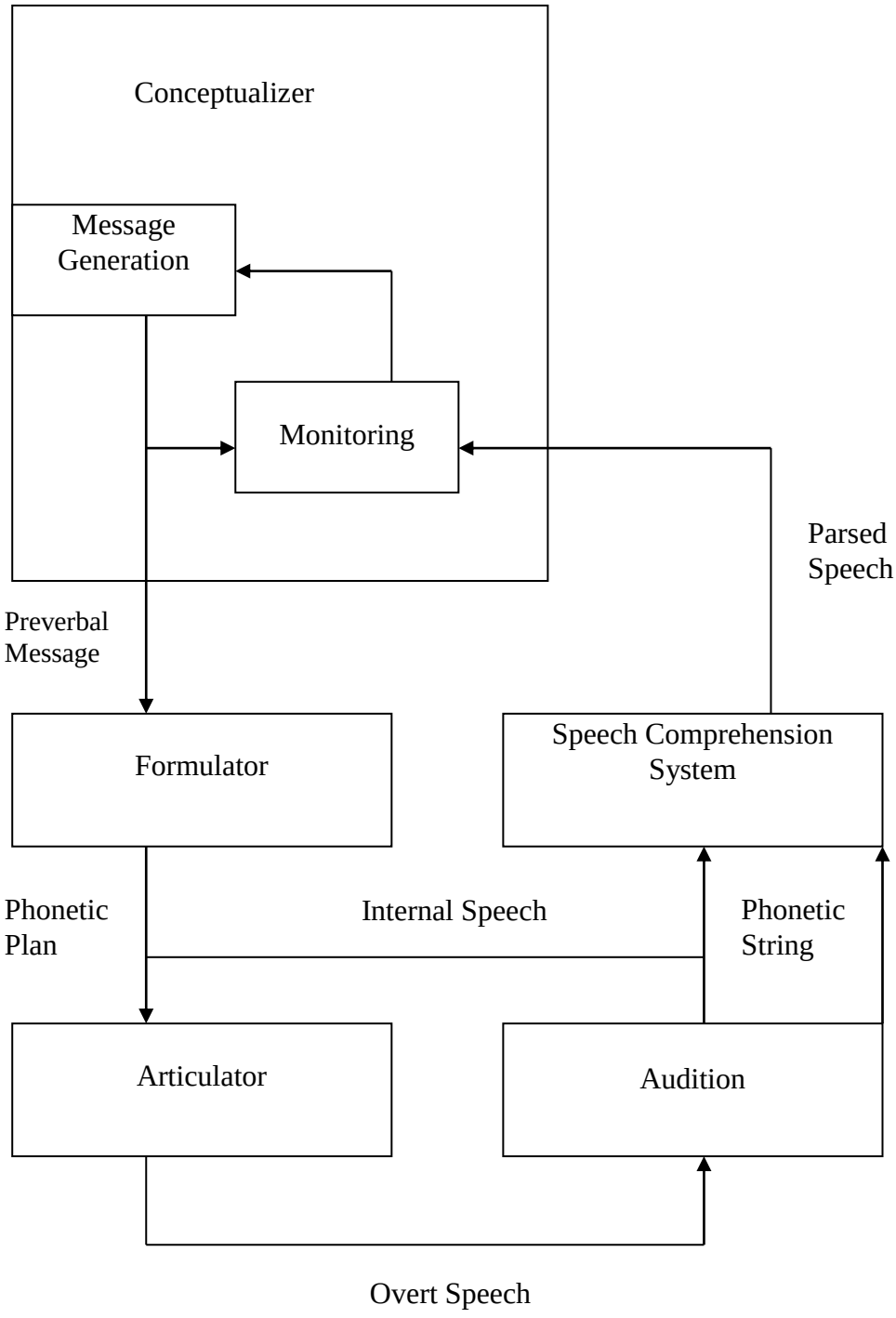
2.1 Introduction

The purpose of this chapter is to discuss the theoretical research relating to L1 and L2 speech production, specifically, to find out how L1 and L2 speech is produced and how L2 speech can lead to L2 development. We begin in 2.2 by looking at Levelt's (1989) psycholinguistic model of L1 speech production. In 2.3 we move onto examine L2 speech by discussing Kormos' (2011) bilingual model of speech production. In 2.4 we then discuss the three aspects of L2 speech to be investigated in this study: fluency, accuracy and complexity and how they can be used as variables for L2 oral development.

2.2 Levelt's (1989) model of L1 speech production

During the past thirty years, numerous psycholinguistic models have emerged to explain how speech is produced (Gilbert, 2007), however, "by far the most influential theory where studies of task planning are concerned is Levelt's (1989) model of speech production" (Ellis, 2005, p. 11). Levelt (1989) reminds us that a speaker is "a highly complex information processor who can, in some rather mysterious way, transform intentions, thoughts, feelings into fluently articulated speech" (p. 1). In order to understand this process, Levelt (1989) devised a psycholinguistic model to explain how speech is produced in the L1 (see figure 1).

Figure 1. Levelt's (1989) model of speech production (cited in Howell, 2004, p. 26)



Levelt's (1989) model comprises of the following four main components which we shall now examine:

- conceptualization
- formulation
- articulation
- self-monitoring

2.2.1 Conceptualization

According to Levelt (1989), the process of speaking begins with the conceptualizer which involves generating the communicative intention of a message and it consists of three sub-stages. The first stage is to decide on the communicative goal of the message. The second stage, referred to as "macro-planning" (p. 107) involves dissecting the goal of the message into sub-goals and then accessing the required speech acts associated with each one from the learner's long-term memory. Speech acts could relate to functions such as requesting or apologizing etc. The third stage, "micro-planning", determines how to express each sub-goal "the information perspective of [an] utterance, its topic, its focus, and the way in which it would attract the addressee's attention" (p. 5). Once the intention of the message has been decided, referred to as a "preverbal message" (p. 9) it is then sent to the formulator to be converted into language.

2.2.2 Formulation

Formulation involves selecting appropriate lexical, phonological or grammatical structures that reflect the content of the preverbal message. This is achieved by accessing "the mental lexicon – the store of information about words in one's language" (Levelt, 1989, p. 6) which is located within the learner's long-term memory. Lexical items are selected for grammatical and phonological encoding by identifying two types

of information within each item: lemmas and lexemes. Lemmas consist of semantic information, syntactic information and may also include morphological information and whereas lexemes consist of phonological information as well as morphological information. Grammatical encoding culminates in a "surface structure – an ordered string of lemmas grouped in phrases and subphrases" (Levelt, 1989, p. 11). The surface structure then enters the phonological encoding of the lexical item's morphological and phonological properties. In order for the planned internal speech to be produced as talk it is sent through the articulation process.

2.2.3 Articulation

Articulation involves the pronunciation and intonation of speech. In order for articulation to take place, the internal speech is first stored in "the Articulatory Buffer. The Articulator retrieves successive chucks of internal speech from this buffer and unfolds them for execution" (Levelt, 1989, p. 13). Transferring internal speech into talk involves using "the motor control of the articulatory organs; in English the lips, tongue, teeth, alveolar palate, velum, glottis, mouth cavity and breath" (Bygate, 2001a, p. 16). This process produces talk, referred to by Levelt (1989) as "overt speech" (p. 13).

2.2.4 Self-Monitoring

All messages, both internal and overt can be stored in the learner's short-term memory (or working memory) where they are checked for errors by a self-monitoring system (Levelt, 1989). Self-monitoring has access to the lexicon so it can recognize words and it enables the speaker to monitor speech in various ways. For example, the intention of messages can be checked during conceptualization, and internal speech can also be monitored before it reaches articulation (Levelt, 1989).

2.2.5 Controlled and automatic processing

In terms of the stages of speech production, conceptualization is a highly controlled conscious process because "communication intentions can vary in finite ways, and for each of these ways the speaker will have to find new means of expression" (Levelt, 1989, p. 21). In other words, there are many ways a message can be conveyed depending on the context therefore conceptualization requires attention in order to generate messages as intended. In addition, self-monitoring is also a controlled process as a speaker is generally aware when making self-corrections to his/her speech. According to Levelt (1989) though, formulation and articulation are carried out automatically which allows L1 speech to be produced without time delays thus facilitating fluent and accurate speech.

2.2.6 Incremental production

Levelt (1989) notes that L1 speech consists of incremental production which means speech is processed in a serial and parallel manner. Serial processing implies that all utterances pass through the same stages i.e. conceptualization, formulation and articulation. Speech production also involves parallel processing as the conceptualizer, formulator and articulator operate simultaneously by attending to different parts of an utterance. This brings us to the end of Levelt's (1989) account of L1 speech production. We now turn to see what similarities or differences exist between L1 and L2 production.

2.3 L2 production

De Bot (1992) points out that "many aspects of speaking are the same for monolingual and bilingual speakers" (p. 2) however, Levelt's (1989) model needs revising in order to take into account certain aspects of L2 production. For example, De Bot (1992) argues that L1 and L2 language processing differs in terms of formulation because L1 language is encoded automatically whereas L2 production involves conscious attention, especially with learners of limited L2 ability who would need time to grammatically encode their communicative intentions. Consequently, Kormos (2011) devised a

bilingual model of oral production which is adapted from Levelt's (1989) model in order to account for how L2 speech is produced.

2.3.1 Kormos' (2011) bilingual model of speech production

Kormos' (2011) bilingual model of speech production is similar to Levelt's (1989) L1 model in that it consists of the same main components: conceptualization, formulation, articulation and self-monitoring, therefore we will briefly review them. First is conceptualization which involves "activating the relative concepts to be encoded" (Kormos, 2011, p. 42). In other words, planning the goal of a message, referred to as "macro-planning" (p. 44) and then deciding on the language perspectives to express it, known as micro planning. This initial planning of a message is not yet linguistic and is also referred to as the preverbal plan. Second is formulation which concerns the lexical, grammatical and phonological encoding of the preverbal plan. The specifications of the plan activate the required lexical items within a learner's mental lexicon. Syntactic encoding begins with the activation of the appropriate lemma, followed by encoding of phrases and clauses. Third is articulation which receives and executes the intended message as spoken language. Finally, there is a self-monitoring component which checks each of the above stages for errors as speech is generated and processed.

Kormos' (2011) bilingual model differs from Levelt's (1989) L1 model in terms of how conceptualization, formulation and articulation are processed in the L2. Regarding L1 speech, grammatical and phonological rules are automatized and assumed to be embedded within the formulator (Kormos, 2011). In terms of L2 production, "rules are not automatic and are assumed to be stored in the form of declarative knowledge" (p. 42). Declarative knowledge refers to factual knowledge, and in the case of language, it refers to knowledge about language including the underlying grammatical rules of the learner's L2 system. For L2 production, Kormos' (2011) bilingual model claims that information is accessed and retrieved through a specific declarative memory store located within the learner's long-term memory. Sub-section 3.2.4 discusses declarative knowledge in more detail.

As we saw in 2.2.6, L1 speech involves incremental production which enables all three stages to operate in parallel with the capacity to produce L1 speech with no time delays (Levelt, 1989). In terms of L2 production, learners typically do not have

automatized knowledge, and so it takes more time to encode messages particularly with words which are rarely used (Samuda & Bygate, 2005). Consequently, Kormos' (2011) bilingual model claims formulation and even articulation may require conscious attention for lower proficiency learners. However, as the human processing system is limited i.e. "we cannot pay attention to an unlimited number of things simultaneously" (Kormos, 2011, p. 51) L2 incremental production i.e. parallel processing of conceptualizaton, formulation and articulation may only be possible with advanced learners. In other words, conscious attention to conceptualisation, formulation and articulation prevents parallel processing from occurring as lower-level learners would be unable to simultaneously attend to all three components but rather attend to each one separately, thus "encoding can only work serially" (p. 41). As a result, this would have adverse affects during online communication as learners would struggle attending to all three stages under the limited time constraints of everyday interaction thus forcing them to trade-off attention between conceptualisation, formulation and articulation. For example, a speaker may be more interested in expressing what they want to say rather than how to say it, therefore focusing on more on conceptualisation i.e. the message content as opposed to formulation and the language required. Consequently, the speaker may be able to convey the meaning of an utterance but it may contain lots of errors. L2 communication breakdowns can therefore occur when a learner has to process or respond in the L2 under the normal time constraints of everyday speech.

As we will see in 3.3.5, numerous studies such as Foster & Skehan (1996); Sangarun (2005) have shown that allocating planning time helps free-up learner's attention to conceptualization, formulation and articulation resulting in improved L2 performance. In addition, to combat the cognitive effort involved in attending to conceptualization, formulation and articulation during online communication, the L2 learner can also rely on formulaic language which the next section will now discuss.

2.3.2 Formulaic Language

Kormos (2011) informs us that during conceptualisation, "not every instance of language is creatively constructed. In fact, the majority of our utterances are combinations of memorized phrases, clauses and sentences, which together are called formulaic language" (p. 46). For native speakers, it typically consists of communicative

functions such as apologizing or requesting and is initiated in conceptualization as ‘chunks’ that contain multiple concepts which activate subsequent linguistic chunks stored in the lexicon as one lemma. For example, the words ‘good’ and ‘morning’ are each stored separately but there is also an additional unit that combines both of them together into one chunk as “good morning” (Kormos, 2011, p. 46). This chunk will then be activated and retrieved when the context is called upon. Formulaic sequences can therefore enhance the speed of L2 production as conceptual chunks can activate matching linguistic chunks for encoding which enables utterances to be “produced faster and with less conscious effort than creatively-constructed elements of the message” (Kormos, 2011, p. 46).

Sinclair (1991) coined the term ‘the idiom principle’ in relation to L2 learners’ use of formulaic language which states:

> a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. (p. 110)

Thus formulaic sequences assist L2 production due to the availability of pre-constructed phrases that can be produced with less effort and attention than creatively constructed messages enabling the learner to produce more fluent speech. Hakuta (1976) and Krashin & Scarcella (1978) (cited in Ellis, 2008) identified two types of formulaic language: routines which refer to a complete phrase memorized as a chunk, for example “I don’t know” (p. 75) and patterns which are partially memorized and have open slots to be filled during communication, for example “Can I have a…..?” (p. 75). As we shall see in 8.3, the use of planning time enables learners to memorize formulaic patterns which lead to improvements in L2 oral performance.

2.3.3 Summary: models of L1 and L2 speech production

This section began by describing the process of L1 speech production according to Levelt's (1989) model which involves four main stages; conceptualisation, formulation, articulation and self-monitoring. We saw how the conceptualizer plans the content of messages, the formulator provides the language to encode it, the articulator produces speech, and self-monitoring checks for errors. During this process, the speaker has control over the conceptualizer and monitoring whilst formulation and articulation are carried out automatically. Finally, L1 production is performed incrementally with conceptualisation, formulation and articulation working simultaneously on different parts of a message. We then saw how L2 speech production differs from the L1 in the following way:

- Use of a declarative memory store
- No automatic processing (all controlled)
- No parallel processing (serial only)

We saw how conscious attention towards L2 speech places pressure on learners' working memory in order to process information during real-time communication which results in learners trading-off attention between conceptualisation, formulation and articulation which has detrimental effects on fluency or accuracy. Finally, we saw how learners can ease the pressure of working memory by using formulaic language which allows speech to be produced quicker and with less effort.

2.4 Aspects of L2 oral production: fluency, accuracy and complexity

As we saw in 2.3.1, it is difficult for L2 learners to process formulation and articulation automatically. So how can the automation of L2 speech be developed? Complexity, accuracy and fluency (CAF) are three important aspects of L2 speech which have been measured and used extensively within second language acquisition (SLA) research during the past twenty years as a means to assess L2 oral performance, proficiency and acquisition (Housen et al., 2012). The problem however, is that "many L2 studies that investigate CAF either do not explicitly define what they mean by these terms, or when

they do, they do so in rather general and vague terms" (p. 3). This has resulted in numerous definitions being put forward to describe the constructs and is one of the reasons why there have been inconsistent results with CAF measures in terms of SLA research (as we will see in 2.4.6). This section begins by reviewing and critiquing various definitions that concern fluency before outlining the definition chosen for the present study. We then provide an additional review of the measures related to fluency. These steps are then repeated for defining accuracy and complexity. Finally, in 2.4.7 we see how CAF can be used as indicators of L2 oral development.

2.4.1 Fluency

Fluency, due to its multifaceted nature, is a difficult and problematic term to define. Various definitions have been used to describe the construct. One of the earliest studies to investigate fluency was Fillmore (1979) who outlined four ways in which a person could be considered fluent:

1. Talking without using many pauses
2. Talking in a coherent manner
3. The ability to talk on a wide range of topics
4. The ability to use language in a creative manner

Although the above points attempt to describe the multidimensional nature of fluency, for example, the ability to speak without pausing, or having the ability to speak about different topics, it is unclear which of them relates more towards L2 fluency. In other words, is talking with fewer pauses a more accurate indication of L2 fluency compared to being able to talk coherently? Furthermore, the above points are vague and open to interpretation. For example, 'talking coherently' could relate to the content of speech that is considered logical or it could relate to how the interlocutor is speaking in terms of pronunciation and articulation, or both content and articulation. Talking without using many pauses is also problematic as it can be perfectly natural to pause a lot in certain contexts, for example, during a group discussion. In addition, a person could

pause a lot to think of what to say but still produce short bursts of speech that would be considered fluent.

Ellis & Barkhuizen (2005) provide a narrower definition as "the production of language in real time without undue pausing or hesitation" (p. 139). Thus, fluency can be seen as the ability to speak in the L2 under natural speaking conditions which for L2 learners generally relates to the ability to speak spontaneously without relying on planning time in order to draw comparisons with the 'automatic' processing conditions of native speakers. This is an important consideration for the present study as the ability to produce L2 speech during unplanned conditions can serve as an indication of acquired knowledge which will be discussed in more detail in 3.5.2. Ellis & Barkhuizen's (2005) definition also involves being able to speak without undue pausing. As we know, for native speakers it is perfectly natural to pause when speaking but for L2 speakers, too many pauses or hesitations could imply difficulty using the L2 and hence a lack of fluency. On the other hand, it might not, for example, what constitutes an undue pause? How can we distinguish undue pauses and hesitations with L2 fluency from other personal and social factors that cause pausing but are not related to L2 proficiency? A learner may have fluent command of the L2 but pauses a lot during a performance because he/she might be feeling tired or shy. Krashen and Terrell's (1983) affective filter theory claims that language learning is most successful when learners have low affective filters and are emotionally stable. In other words, acquisition and performance occurs best when learners have low levels of anxiety etc. Fluency could there be disrupted by stress which could mask a learner's proficiency of the L2.

Tavakoli & Skehan (2005) suggest that undue pausing can be assessed by using breakdown fluency measures that involve the number and length of pauses. "There is, though, some disagreement regarding the minimum length for a pause to be counted as a pause, with proposals as low as .25 of a second" (p. 254). Other studies such as Freed (2000) (cited in Tavakoli & Skehan, 2005) measured fluency using unfilled pauses that were classified as disfluent if they lasted 0.4 seconds or longer. It seems disagreements would always exist regarding what constitutes the minimum length of a pause depending on the context.

According to Skehan & Foster (1999), fluency is "the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems" (p. 96). This is a similar definition to Ellis & Barkuizen (2005) in that fluency relates to the

ability to produce language spontaneously but with an emphasis on expressing meaning as opposed to concentrating on language form. Although it could be argued that a fluent L2 speaker has the capacity to emphasize meaning *and* form, for example, expressing an opinion without making a mistake. This leads us onto Lennon's (2000) definition of fluency which represents "the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (p. 26). This definition goes beyond the ability to speak under natural time constraints to also include accuracy as an indicator of fluency. Lennon (2000) adds that fluency is not only restricted to the productive skill of speaking but it also applies to writing as well as the receptive skills of listening and reading. Consequently, those who cannot understand speech do not have fluent receptive skills.

As we can see, fluency is a multidimensional construct (Housen et al., 2012), and as a result, it is difficult to define. After reviewing various definitions and discussing the limitations of them, for the purpose of this study, it is perhaps best to synthesize the above terms into a working definition. Although Kormos & Denes (2004) point out that in terms of L2 oral fluency, Lennon's (2000) definition successfully combines the strengths of previous ones. Consequently, this study shall rely on Lennon's (2000) definition but with the omission of 'accuracy' as although it can be seen as an aspect of fluency, for the purpose of this thesis, accuracy is used as a separate construct described in 2.4.3. Thus Lennon's (2000) definition is adapted as follows, "the rapid, smooth, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (p. 26). This definition combines several aspects of fluency and each one will now be summarized in relation to the present study.

1. 'Rapid' concerns the speed of L2 delivery i.e. the ability produce speech in real-time speaking conditions that do not involve conscious planning time as outlined in Ellis & Barkhuizen (2005).
2. 'Smooth' relates to the use of formulaic language which was discussed in 2.3.2 and allows the learner to produce 'chunks' of language such as communicative functions which are easier and faster to produce than individual units (Kormos, 2011).

3. ‘Lucid’ relates to the ability to produce L2 speech which is understandable to others. For the purpose of the present study, lucid speech relates to pronunciation and intonation. Other factors that could relate to lucid speech such as grammatical accuracy will be analyzed separately in 2.4.3 when we discuss accuracy as a separate construct.
4. ‘Efficient translation of thought or communicative intention into language’ conveniently refers to the process of L2 speech production outlined in Kormos’ (2011) model in 2.3.1. We know that learners usually do not have automatized L2 knowledge and so it may take time to encode preverbal messages. However, fluency is the ability to process conceptualisation, formulation and articulation efficiently in order to produce L2 speech without time delays.
5. ‘temporal constraints of on-line processing’ refers to the ability to produce language under the natural time constraints of everyday speech which as we discussed in 2.3.1 typically does not involve planning time. Thus L2 learners are required to produce speech ‘online’ which as mentioned in point 4 requires efficient processing of communicative messages.

Although this definition may still be limited in terms additional aspects of oral fluency that may not be accounted for, it does appear to cover all the main areas to be considered for the present study and which can be measured. The following sub-section discusses fluency measures in more detail before briefly outlining the measures used for the present study that attempt to reflect our definition.

2.4.2 Fluency measures

Due to the problematic nature of defining fluency, operationalising the construct is also a complex matter. Not surprisingly then, fluency has been measured in different ways. Tavakoli & Skehan (2005) outline three main types of measures: temporal or speech rate measures, for example, number of syllables per minute. Fluency breakdown measures discussed above such as number of pauses. Finally, measures relating to

repair fluency, for example, number of false starts, or repetitions. Table 1 provides an illustration of the measures used in task planning research.

Table 1. *Measures for assessing fluency (adapted from Ellis, 2005, p. 32)*

Type of Measure		Description	Planning Study
Fluency	Speech rate	The number of syllables produced per minute of speech	Kawauchi (2005), Sanguran (2005), Yuan & Ellis (2003), Mochizuki & Ortega (2008), Mehnert (1998), Ortega (1999), Gilabert (2007b)
	Breakdown fluency	The ratio between number of words reformulated and total words produced	Yuan & Ellis (2003), Skehan & Foster (1996, 1999)
		Total silence	Skehan & Foster (1996,1999)
		Number of pauses greater than 1 second	Tavokoli & Skehan (2005)
		Number of filled pauses	Mehnert (1998)
	Repair fluency	Number of repetitions	Kawauchi (2005)

As with defining fluency, measures relating to the construct also appear to have weaknesses. For example, speech rate measures such as 'syllables per minute' could prove problematic as an indication of L2 proficiency because syllables could include L1 use. Furthermore, a learner could repeatedly use the same words again and again thus sounding incoherent yet would appear to be fluent due to the amount of syllables produced. In addition, breakdown fluency measures as discussed in the last sub-section have weaknesses as pausing for more than one second may not reflect disfluency, for example narrating a story often requires pausing as a means to signal a change of topic. As a result of these issues, as well as considering the definition of the present study, the following measure was used 'pruned speech rate' which relates to "the average number of syllables produced per minute of pruned speech, i.e. speech from which repetitions, false starts and other performance features

have been excluded" (Levkina & Gilabert, 2010, p. 182). In 5.7.4, we describe in detail the justifications for choosing this measure but for the purpose of this sub-section the following brief explanation will be provided. A pruned speech rate measure was considered to be a suitable reflection of our working definition for the following reasons:

1. Syllables per minute is a speech rate measure which "deals with the speed with which language is produced" (Tavakoli & Skehan, 2005, p. 254). As a result, it represents the 'rapid' aspect of our definition in point 1 as we can measure how fast learners produce L2 speech. This measure could also calculate the 'smooth' use of formulaic language in point 2 by counting the syllables of 'chunked language' produced per minute.
2. Excluding "repetitions, false starts and other performance features" (Levkina & Gilabert, 2010, p. 182) helps to ensure that the language we are measuring represents point 4 'efficient translation of thought or communicative intention into language' as we are only interested in analysing L2 language intended by the speaker. Thus, for the purpose of this study, 'other performance features' relates to irrelevant language such as L1 use, self-corrections and incomprehensible language. In doing so, the elimination of these features would reflect 'lucid' language in point 3 that is understandable.

In order to measure fluency in terms of point 5 of our definition: 'under the temporal constraints of on-line processing' requires a measurement of assessment as opposed to a measurement of speech production. In other words, in order to assess learners' fluency under the constraints of on-line processing would require conditions that do not allow planning time. For example, asking a learner to narrate a story without having the opportunity to plan what to say would require the learner to produce speech spontaneously. Consequently, the present study designed oral narrative tests that did not allow planning time in order to test learners' fluency under the constraints of on-line processing. We will discuss the testing measures used for the present study in 5.3.1 however, for now after confirming our definition for fluency and the measure used to assess it we now turn our attention towards accuracy.

2.4.3 Accuracy

Accuracy, like fluency, is problematic to define particularly when attempting to distinguish it from fluency as both constructs can be considered to be aspects of the other in terms of indicators of L2 oral proficiency, as we saw in Lennon's (2000) definition. As a result, this sub-section shall review various definitions of accuracy before highlighting the weaknesses of them and then consolidating a definition for the purpose of this study. Yuan & Ellis (2003) define accuracy as "the extent to which the language produced conforms to target language norms" (p. 2). This means the level to which speaking in the L2 meets the standards of a native speaker in terms of the amount of mistakes or errors made. Brumfit (1984) was one of the first SLA researchers who chose not to differentiate fluency and accuracy on linguistic terms but rather to distinguish them within pedagogic contexts:

> accuracy and fluency is essentially a methodological distinction, rather than one in psychology or linguistics. That is to say, it is a distinction which may have values to teachers in decision making about the content of lessons and the distribution of time between various types of activity. (p. 52)

Byrne (1987) contrasts accuracy activities with fluency activities. The former is "to make sure that students get enough practice in a particular point of grammar or vocabulary or pronunciation" (p. 7). Accuracy exercises are used to draw students' attention to specific areas of language form in order to help them speak correctly. Whereas the latter allow "your students opportunities to *use* the language they have learnt: to use it freely, even if they make mistakes". In other words, fluency activities focus on meaning and allow students to use language for communication. However, as this study is interested in analysing accuracy in terms of L2 oral performance, a linguistic definition is preferred over a pedagogical one. Housen et al. (2012) echo Yuan & Ellis' (2003) definition by claiming that:

> accuracy (or *correctness*) in essence refers to the extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm (i.e. usually the native speaker). Thus, deviations from targetlike performance would be considered errors. (2012, p. 4).

Although Pallotti (2009) rightly points out that caution must taken when applying accuracy measures that relate to errors as "one can have perfectly accurate but communicatively inadequate messages (*colorless green ideas . . .*)" (p. 592) which of course can lead to misunderstandings. Furthermore, it is unclear what deviations from targetlike norms relate to. Does it concern grammatical errors or errors related to pronunciation? If it is the latter, although pronunciation is clearly important in terms of oral performance, would it be unfair to claim a Japanese intermediate learner's L2 speech contains errors even though it may be grammatically correct and understandable but their pronunciation does not reflect *native* speaker norms? After all, it is well documented that English and Japanese languages contain phonological differences that result in pronunciation problems for Japanese learners of English, particularly in the use of 'r' and 'l' consonants (Ohata, 2004). In addition, pronunciation is problematic because we have referred to it as an aspect of fluency in 2.4.1 which represents 'lucid' speech that is understandable to others.

To counter the weaknesses of accuracy definitions that relate to errors, Housen et al. (2012) commented that "the A in CAF be interpreted not only as accuracy in the narrowest sense of the term but also as *appropriateness* and *acceptability*" (p. 4). In doing so, this eliminates Pallotti's (2009) criticism of having grammatically accurate utterances that are communicatively inadequate as they would be considered inappropriate. Thus given the limitations of earlier definitions we shall rely on the definition of Housen et al. (2012) as "the extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm (i.e. usually the native speaker). Thus, deviations from targetlike performance would be considered errors" (p. 4). In addition, accuracy also relates to "*appropriateness* and *acceptability*" (p. 4). As the purpose of this study is to track learners' development of specific grammatical features, specifically English relative clauses and 3rd person singular and plural (which will be discussed in more detail in 3.5.3) deviations from targetlike performance relate solely to grammatical errors and communicatively inadequate use of the targeted forms. In other words, accuracy relates to grammatical errors concerning relative clauses and 3rd person singular and plural as well as communicatively adequate use of the forms. Let us examine each of these two factors in turn. In terms of grammatical accuracy of the targeted forms, targetlike performance could be:

'He thinks that he likes the dog which has long ears'

In this case, there are two instances of 3rd person singular *'he thinks'* and *'he likes'* that accompany the relative clause *'which has long ears'*. Thus, targetlike supplanted of the targeted forms consist of no grammatical errors. In terms of the relative clause, this involves no grammatical errors within the clause such as verb tense and the use of articles. Consequently, deviations from targetlike performance would involve grammatical errors relating to the targeted forms, for example, verb tense within the relative clause:

'He likes the dog which have long ears'

Deviations from targetlike performance concerning 3rd person singular would involve grammatical errors relating to the use of the form, for example, incorrect subject-verb agreement:

'He think he likes…..'

In the case of communicatively adequate use of the targeted forms, accuracy would relate to producing the targeted forms in the context in which they should be used. For the purpose of the present study, this involved using the forms to narrate a story. Thus, if a learner produced a relative clause that was grammatically correct but it did not reflect the context of the storyline it would be considered inaccurate. For example, describing a picture of a dog with long ears but commenting "the cat *which has short ears*". In addition, communicatively adequate use of the forms relates to over-use of the forms. For example, repeating the same relative clause again and again when narrating a story would be considered communicatively inadequate, and classified as inaccurate, for example, as in the underlined structure:

'he likes the dog <u>*which has long hair*</u> *which has long hair'.*

In 5.7.2 we examine measures relating to grammatical accuracy and communicatively adequate use of the targeted forms for the main study in more detail. To conclude, although our definition of accuracy may be limited, for the purposes of this study it does cover the main areas we wish to consider. In the next sub-section, we

examine the measures used for accuracy before confirming the measure used in the present study.

2.4.4 Accuracy measures

As with fluency, previous studies have measured accuracy in various ways (see table 2). Some studies such as Skehan & Foster (1996) elected to use general measures for accuracy that relate to 'error-free clauses' whilst other studies such as Kawauchi (2005) used specific measures to investigate how accurate certain grammatical forms are used during production e.g. past-tense markers.

Table 2. *Measures for assessing accuracy (adapted from Ellis, 2005, p. 32)*

Type of Measure		Description	Study
Accuracy	Overall grammatical accuracy	Error-free clauses	Yuan & Ellis (2003), Mehnert (1998),
		Error-free clauses of different lengths	Skehan & Foster (1996, 1999)
		Number of errors per 100 words	Sangaran (2005), Mehnert (1998)
		Correct verb forms	Yuan & Ellis (2003)
	System-based grammatical accuracy	Past-tense markers	Kawauchi (2005)
		Quality of relative clauses	Mochizuki & Ortega (2008)
		Discourse markers	Williams (1992)

In line with the fluency measures described in 2.4.2, there appears to be certain issues with the above measures as indicators of oral accuracy. For example, as we mentioned

in the previous sub-section, it is unclear whether 'error-free clauses' relate to grammatical correctness or errors in pronunciation. If it is both, it would surely prove to be an unrealistic measure of accuracy for lower proficiency learners under testing conditions. Furthermore, as we discussed in our definition, accuracy measures are limited unless they specify clearly the level of appropriateness and acceptability in terms of the context in which the language is used. For example, if we take the measure 'number of errors per 100 words', a learner could repeat the same phrase again and again during a narration without producing errors. At the end of the narration, the learner could appear to be an accurate speaker despite only using a limited amount of vocabulary. Thus although these accuracy measures account for grammatical accuracy they do not seem to consider the appropriateness of accurate language within a social context. In 5.7.2, we return to these issues in order to justify the accuracy measure of the present study however, in order for clarity, a brief description of the measure used is now provided. A rating scale measure was chosen to assess learners' grammatical accuracy concerning the production of targeted linguistic forms of the present study: English relative clauses and 3rd person singular and plural. The learners were expected to produce the forms whilst performing a story-telling narrative that was designed to elicit seven obligatory instances of the forms. In terms of communicative adequacy, learners' accuracy was also dependent upon using the targeted forms appropriately in relation to the storyline. We will discuss how this was achieved in 5.7.2. We now turn our attention towards the final aspect of L2 speech: complexity.

2.4.5 Complexity

Complexity, like fluency and accuracy is mutli-dimensional in nature however complexity in particular, has drawn the most controversy in the field (Bulte & Housen, 2012; Norris & Oretga, 2009). The term complexity has been used interchangeably within SLA literature to refer to different concepts of complexity. Even within the scope of this study it will be used to refer to cognitive complexity and linguistic complexity. The former relates to the difficulty with which learners process language under different conditions (Housen et al., 2012), for example, performing a complex task is generally considered to be more cognitively demanding (more difficult) than a simple task. Cognitive complexity will be discussed in more detail in 3.4.3. Linguistic complexity,

on the other hand, is a component of cognitive complexity but it refers to the linguistic features of language considered to be complex in terms of construction or rule. Ellis & Barkhuizen (2005) define linguistic complexity as "the extent to which learners produce elaborated language" (p. 139). This involves the ability to produce more complex forms of language than what is usually expected of the learner's proficiency. In order to achieve this, complexity requires the learner to have sufficient linguistic knowledge of the L2 and is therefore difficult for beginner level learners to produce. Skekan & Foster (1999) define complexity as:

> the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively. This may also involve a greater willingness to take risks, and use fewer controlled language subsystems. This area is also likely to correlate with a greater likelihood of restructuring, that is, change in the interlanguage system. (p. 97)

Complexity therefore involves a degree of risk, as the production of elaborate language may be more difficult to control and susceptible to errors. Skehan (1998) argues that complexity may be a more important aspect of L2 speech compared to fluency or accuracy as it requires learners to "use language closer to the 'cutting edge' of their language development" (p. 69). In other words, attempting to produce more complex speech in the form of *new* language, as opposed to relying on the use of already acquired forms will help to develop learners' L2 oral proficiency. Consequently, developments in complexity can reflect improvements in learners' interlanguage (i.e. their current L2 knowledge), in that the more input or knowledge that is acquired, the more linguistically complex a learner's L2 output becomes.

As we have seen so far, complexity is a mutlifaceted construct. Even when narrowed down to linguistic performance, it can still refer to lexical, morphological, syntactic, or phonological complexity. As a result, complexity is difficult to define. According to Bulte & Housen (2012), "many L2 studies that investigate 'complexity' either do not define what they mean by this term, or when they do, they do so in general, vague or even circular terms." (p. 22). For example, as we saw with Skehan & Foster's (1999) definition: "the capacity to use more advanced language…." (p. 97) is extremely vague as it is unclear what is meant by term 'advanced language'. Does it relate to proficiency or certain grammatical features? An equal lack of clarity applies to Ellis & Barkhuizen's (2005) definition of 'elaborated language'. Does this relate to fluency,

accuracy, pronunciation, or specific linguistic features? Not surprisingly then, SLA studies have provided mixed results concerning the variable as an aspect of oral production. Recently, Bulte & Housen (2012) attempted to provide a more specific definition as:

> a language feature or system of features is seen as complex if it is somehow costly or taxing for language users and learners, particularly in terms of the mental effort or resources that they have to invest in processing or internalizing the feature(s). (p. 23)

For example, Bulte & Housen (2012) note that English relative clauses have been shown to be more difficult to produce and tend to be acquired later than other linguistic forms such as coordinate structures. Reasons for this may be learner dependent, for example, the complexity of linguistic features may depend on factors such as the learners' L1 background. As we shall see in 3.5.4, English relative clauses are known for their difficulty in production with Japanese learners because of differences that exist between English and Japanese versions of the form. Thus English relative clauses could be considered a complex linguistic feature for Japanese learners. As the purpose of this study is to investigate L2 development of grammatical structures, specifically English relative clauses, we shall focus solely on syntactic complexity, also known as grammatical or structural complexity, and rely on Bulte & Housen's (2012) definition as:

> a language feature or system of features is seen as complex if it is somehow costly or taxing for language users and learners, particularly in terms of the mental effort or resources that they have to invest in processing or internalizing the feature(s). (p. 23)

Although this definition is limited in its scope by referring to complexity as linguistic features known their cognitive difficulty in use or acquisition, it does fit the purpose of this study which is to highlight the use of a linguistic feature known for its difficulty in L2 oral production. In 3.5.4 we will find out in more detail why English relative clauses are considered to be a cognitively difficult aspect of English grammar for Japanese learners. However, for now, it is important to identify a measure of complexity that reflects our definition which the following sub-section shall now discuss.

2.4.6 Complexity measures

According to Yuan & Ellis (2003) "measures of complexity are generally based on the extent to which subordination is evident" (p. 2). In other words, assessing learners' use of multiple clauses, for example dependant clauses which are compared against measuring units such as 't-units' or 'c-units'. Mochizuki & Ortega (2008) define a t-unit as an utterance that contains "a main clause with or without subordinate clauses" (p. 23). These measures relate more towards syntactic complexity however, lexical complexity can also be measured in a variety of ways (see table 3).

Table 3. Measures for assessing complexity (adapted from Ellis, 2005, p. 32)

Type of Measure		Description	Study
Complexity	Syntactic complexity	Ratio of clauses to some general unit (e.g. t-units, c-units or AS-units)	Kawauchi (2005), Yuan & Ellis (2003), Sangaran (2005), Williams (1992), Skehan & Foster (1996, 1999)
		Length of unit (e.g. t-unit)	Kawauchi (2005)
		Number of relative clauses per t-unit	Mochizuki & Ortega (2008)
		Number of subordinate clauses	Kawauchi (2005), Mehnert (1998)
	Complex grammatical structures	Use of comparatives and conditionals	Kawauchi (2005)
	Syntactic variety	Total number of different grammatical verb forms used in the task	Yuan & Ellis (2003)
	Lexical variety	Mean segmental type/token ratio Guiraud's index of lexical richness	Yuan & Ellis (2003) Gilabert (2007b)

Due to the multifaceted nature of complexity, Bulte & Housen (2012) note that "none of the complexity measures employed or recommended in the L2 research is unproblematic" (p. 40), as a result, a number of criticisms have been aimed at the validity of the measures used to assess the construct. For example, regarding syntactic complexity, Bulte & Housen (2012) point out that subordination measures such as 'clauses per t-unit, or 'relative clauses per t-unit' are considered to be hybrid measures in that they are supposed to capture specific aspects of subordination as well as 'difficulty'. Subordination measures have

been used as an indication of syntactic complexity because they are known for their cognitive difficulty in oral production (as we shall see in 3.5.4 when we analyze relative clauses). However, Pallotti (2009) questions whether the relationship between *difficulty* and *structural complexity* remains the same for *all* forms of syntactic structures. Perhaps certain types of clauses, for example dependent clauses or relative clauses become easy to produce once L2 learners reach a certain proficiency, or perhaps certain forms of L2 subordination are similar to the learner's L1 equivalent, and therefore may not accurately reflect complex language. Furthermore, given the multidimensional nature of linguistic complexity, as well as the various measures that accompany it, "one might wonder whether it is appropriate to consider all these aspects as dimensions of the same construct or as different constructs altogether" (Pallotti, 2009, p. 593).

Although these issues will be discussed further in 5.7.3 when we justify the measure used to assess syntactic complexity for the present study, a brief description of the measure used 'relative clauses per AS-unit' will now be provided. An AS-unit is similar to a t-unit in that it comprises of a main clause and or any subordinate clauses attached with it (Foster, Tonkyn & Wigglesworth, 2000). Relative clauses were chosen because the grammatical feature is known for its cognitive difficulty in oral production with Japanese learners of intermediate proficiency. As the purpose of this study is to develop Japanese intermediate-level learners' use of the form, 'relative clauses per AS-unit' was considered to be a suitable measure of complexity that would reflect our definition of a linguistic feature that is cognitively challenging for learners to use. In terms of production, non-target-like use of relative clauses was accepted provided the participant used a relative pronoun. However, in line with our accuracy measure, repeated relative clauses were excluded from the analysis in order to prevent over-use of the form, for example, as in the underlined structure 'he likes the dog *which has black hair, which has black hair*'.

2.4.7 Complexity, accuracy and fluency as variables for L2 oral development

As we will see in 3.3.5, CAF measures have been used widely to assess L2 *performance* at certain points in time, however CAF has also been used to imply *developmental* changes in a learner's internal L2 system over time in three ways (Housen et al., 2012). For example:

> (i) internalization of new L2 elements (or greater *complexity*, as more elaborate and more sophisticated L2 knowledge systems are developed); (ii) modification of L2 knowledge (as learners restructure and fine-tune their L2 knowledge, including the deviant or non-targetlike aspects of their interlanguage (IL) so that they become not only more complex but also more *accurate* L2 users); (iii) consolidation and proceduralization of L2 knowledge (i.e. higher *fluency*, through routinisation, lexicalisation and automatisation of L2 elements leading to greater performance control over the L2 system. (p. 3)

In other words, improvements in complexity have been associated with developments in L2 knowledge as a result of learning new grammatical structures, rules and vocabulary. Improvements in accuracy represent restructuring of a learner's internal L2 system in order to meet targetlike performance and eliminate errors during oral production. Finally, improvements in fluency indicate the learner's ability to access their L2 resources to produce language with reduced time delays thus showing greater control of their L2 system. In 3.4.3 we will discuss a theoretically proven model, referred to as the Cognition Hypothesis (Robinson, 2003) which provides a guideline for sequencing tasks in order to facilitate L2 oral development in terms of CAF.

2.5 Chapter conclusion

This chapter has explained the mechanisms involved in L1 and L2 speech production and the implications for L2 oral development. We began by describing Levelt's (1989) model of L1 speech production before moving onto discuss Kormos' (2011) bilingual model of speech production and the role of formulaic language. We then outlined three aspects of L2 speech to be investigated in this study:

- Fluency: the rapid, smooth, lucid, efficient production of L2 language during online communication
- Accuracy: grammatical use of English relative clauses and 3rd person singular and plural in relation to target-like norms including appropriateness

- Syntactic complexity: use of relative clauses which are known for their difficulty in oral production with Japanese learners

We have seen how all three aspects of L2 speech are multidimensional in nature which has led to various definitions and measures used to capture each variable. Nevertheless, these aspects have been widely used in the literature as indicators of L2 oral performance and they also appear valid indicators of L2 oral development. However, as we will see with task planning research in 3.3.5, CAF variables can also compete with each other during L2 performance resulting in improvements in one measure at the expense of others. In the next chapter we examine the pedagogic tools used to facilitate improvements in fluency, accuracy and complexity.

3 LITERATURE REVIEW: PART TWO

3.1 Introduction

The purpose of this chapter is to discuss theoretical research relating to the tools used to promote L2 oral development, namely, tasks, planning conditions, task sequencing and the conditions used to facilitate L2 development. We begin in 3.2 by defining what we mean by 'task' before moving on to examine focused tasks. In 3.2.2 we look at task-based language teaching (TBLT) followed by an account of Willis' (1996) TBLT framework. 3.2.4 critically reviews this framework and outlines an alternative methodology known as task-supported language teaching (TSLT). In 3.2.5 we discuss issues related to TBLT in Japan then in 3.3 we look at the specific area of task-based research for this study: pre-task planning and its component guided planning. 3.4 examines task sequencing; specifically task repetition and task complexity. 3.5 then outlines the effects of guided planning and task complexity on L2 oral development in terms of fluency, accuracy and complexity before finally moving to discuss research concerning the main linguistic feature of the present study: English relative clauses. 3.6 concludes the chapter and outlines the preliminary research questions of the thesis.

3.2 Tasks

During the past twenty years there has been a considerable amount of research carried out on pedagogic tasks as a means for developing L2 performance (Ellis, 2009a). As a result, numerous definitions have been put forward on what a task is or should be, for example, "activities where the target language is used by the learner for a communicative purpose (goal) in order to achieve an outcome" (Willis, 1996, p. 23) or "an activity which requires learners to use language, with an emphasis on meaning, to attain an objective" (Bygate et al., 2001, p. 11). The general consensus is that tasks allow learners to use their L2 to interact with each other in order to reach a goal. However, these terms can appear vague as a 'goal' or 'objective' could relate to any

type of outcome. As a result, Samuda & Bygate (2008) provide a more detailed definition:

> A task is a holistic activity which engages language use in order to achieve some non-linguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning, through process or product or both. (p. 65)

'Holistic activity' refers to a task in its entirety, it can also imply the use of all four language skills not just one area for example, pronunciation. A non-linguistic outcome implies that the goal of the task is not language focused but that language use is required to complete it. A task attempts to provide a 'linguistic challenge' meaning that learners are required to use their cognitive abilities in order to communicate to solve the problem. Finally, a task can be used in different ways to develop learners' language. Language learning could occur through 'process' for example, providing corrective feedback on language used by learners as they attempt a task. Alternatively, language learning could take place through 'product' which involves feedback on the language used after the task e.g. analysing language used to present on the task's outcome (Samuda & Bygate, 2008).

Bygate et al. (2001) suggest a task can be defined depending on its purpose. As the purpose of this study involves learners using grammatical features to express meaning, we shall rely on Samuda & Bygate's (2008) definition.

3.2.1 Focused tasks

As we have seen, tasks place an emphasis on language use i.e. meaning, however, they can also be used to focus on particular areas of language form whilst engaging in language use. These types of tasks are referred to as 'focused tasks' which "aim to induce learners to process, receptively or productively, some linguistic feature" (Ellis, 2003, p. 16). Focused tasks have two aims; to promote language use and facilitate the use of a targeted language form. For example, as we will see in 3.3.6, a focused task could involve learners using relative clauses as they narrate a story. According to Loschky & Bley-Vroman (1993), a focused task should make the use of a targeted form as essential as possible to ensure learners use it. However, as Ellis (2003) points out "it is not easy to design such tasks" (p. 17). Due to a task's emphasis on language meaning, learners can use whatever language they wish to complete a task and so can quite easily

avoid using specific targeted forms (Seedhouse, 2005). However, 3.3.3 describes the means available for teachers to target specific language use, for example, by providing guiding planning prior to a task which can focus learners' attention on using specific forms during performance.

3.2.2 Task-based language teaching

Advocates of task-based language teaching (TBLT) believe that "tasks are both necessary and sufficient for learning" (Ellis, 2003 p. 28). TBLT involves lessons or entire courses that are based around the use of tasks as a means to develop learners' competency in the L2. TBLT however, involves "a number of rather different approaches to using tasks" (p .31) which we will now examine.

TBLT draws influence from communicative language teaching (CLT) which "aims to develop the ability of learners to use language in real communication" (Ellis, 2003, p. 27). CLT developed during the early 1980's as an alternative teaching approach to more traditional teaching methods such as audiolingualism which focused on the study of language form through use of accuracy activities that involved repetitive drills of language structures. As CLT became more widespread, it became categorized into 'weak' and 'strong' versions (Howatt, 1984, p. 279). The weak version uses language functions such as 'apologizing' or 'inviting' as a means of language instruction. This led to the introduction of notional/functional syllabuses that categorized various communicative functions for students to learn and practice in a communicative context. The strong version of CLT "advances the claim that language is acquired through communication". Under this version, students are not presented with pre-determined language functions which are then practiced, but instead learners experience how to use a language through interaction whilst attention to language form occurs incidentally either during or after communication.

It is worth noting at this point the distinction in SLA literature between a focus on form vs focus on forms. "Focus-on-forms instruction involves the pre-selection of specific features based on a linguistic syllabus and the intensive and systematic treatment of those features" (Ellis, 2008, p. 255). This refers to the weak version of CLT, specifically the use of notional/functional and structural syllables which identifies pre-

determined linguistic features to be instructed systematically. "*Focus on form* refers to how attentional resources are allocated" (Long & Robinson, 1998, p. 23). This typically involves shifting learners' attention to language form during a meaning based lesson in order to facilitate learning, for example, providing corrective feedback on language used during interaction.

So how are tasks and a focus on form implemented within syllabuses? We examine this by looking at the strong version of CLT, referred to as TBLT, which uses tasks as the central means for facilitating language learning. One of the earliest and most famous studies involving TBLT was carried out by Prabhu (1987), who introduced one of the first 'task-based' curriculums in secondary schools throughout southern India from 1979-1985, known as 'The Bangalore Communicational Teaching Project' (CTP). During this period, notional-functional and structural syllabuses were prevalent. However, Prabhu (1987) was dissatisfied with them as they involved explicit instruction of language form which he believed to be ineffective because "linguists' generalisations about language structure are unlikely to match whatever generalisations are involved in the learner's process of grammar construction" (p. 144). Prabhu (1987) argued that a more effective method of teaching would come from a syllabus that contained no prescribed instruction towards language form, and this was used for the CTP. It was referred to as a 'procedural syllabus' and it consisted of sequencing meaning-based activities that contained 'pre-tasks' that would be carried out by the teacher and the class as a whole and then students would work on similar tasks themselves in pairs or groups.

The CTP received positive feedback from evaluation tests however Willis & Willis (2001) remind us that it is extremely difficult to prove the findings of such a project. For example, as Murphy (2013) points out, are we to assume all Prabhu's (1987) learners were content to be taught through a focus on meaning? What would happen if learners asked for explicit grammatical instruction? Were all the non-native teachers willing to adopt a new method of instruction concerning free-language use? These issues outline the difficulty in successfully implementing a task-based syllabus in non-western environments where both non-native teachers and learners may not be accustomed to the methods of instruction. In 3.2.5, these issues surface again with regards to implementing TBLT in Japan.

Ellis (2003) distinguishes TBLT as the design for a curriculum and as a methodology for teaching, in that the former concerns selecting and sequencing tasks for learners to participate in during a course program. The latter however, relates to *how* tasks can be used in the classroom. One particular methodology which is widely referred to in SLA literature is Willis' (1996) TBLT framework which the next sub-section will now discuss.

3.2.3 Willis' (1996) TBLT framework

Willis' (1996) TBLT framework is based around a task consisting of three main stages:

- **Pre-task**: introduction to topic and task, preparation
- **Task cycle**: task performance, planning and report
- **Language focus**: language analysis, practice

The pre-task stage involves the teacher providing instructions about the task and brainstorming any useful vocabulary or phrases that could help learners perform the task. According to Willis (1996), the purpose of the pre-task stage is to activate students' own linguistic resources to prepare them for the task cycle. In the task cycle, learners participate in groups, pairs or individually depending on the task-type. During the task, the learners would use their L2 resources to interact with each other whilst the teacher would observe from a distance and provide assistance if needed. Any linguistic difficulties that a student may face during the task could be over come by the group collectively negotiating the meaning of what the learner wants to say. On completion of the task cycle, each group would collectively prepare a report on their findings and present it to the rest of the class and the teacher would comment on any issues. Finally, in the language focus part of the lesson, students would analyse language forms used during the task. The teacher would use this part of the lesson to allow learners to notice new forms of language and then practice using them by repeating parts of the task.

3.2.4 A critical review of Willis' (1996) framework

Although Willis' (1996) framework looks like an attractive model for instruction, Samuda & Bygate (2008) point out that it has not been implemented within task-based programs to any significant degree, whilst there is a lack of empirical studies that have investigated its desirability as a model of instruction. One of the reasons for this is that it appears to conflict with skill-building models of instruction. Anderson's (2000) skill building theory claims that skill development occurs through transforming declarative knowledge into procedural knowledge. As discussed in 2.3.1 "declarative knowledge is factual. In the case of language it involves explicit knowledge of grammatical rules" (Ellis, 2003, p.145). A learner with declarative knowledge of an L2 may know the grammatical rules but may not have had exposure using the language and therefore may struggle in real-time communication. On the other hand, "procedural knowledge is declarative knowledge that has become fully automatized". In this case, the L2 learner has the ability to use a language proficiently without thinking. The transformation of declarative knowledge into procedural knowledge, referred to as 'proceduralisation' is achieved through practice (this process will be examined in more detail in 3.4.3). Anderson's (2000) skill theory appears inline with traditional approaches to language teaching which are still widely used such as PPP which favours presenting explicit instruction of language rules followed by practice until the rules become automatised (Ellis, 2003). Swan (2005) stresses the importance of practice in converting declarative knowledge into procedural knowledge and he argues that PPP is much more compatible with skill learning theories than TBLT, as the former attends to language and practice after a task.

In terms of proceduralisation, Ellis (2003) argues that "for practice to work it must involve learners producing the target structure in the context of communicative activity" (p. 146). Dekeyser (1998) favours this form of language learning in which linguistic structures are first taught declaratively and then proceduralisaton takes place through communicative practice, which he refers to as "engaging in the target behaviour – or procedure – while temporarily leaning on declarative crutches" (p. 49). Through this form of practice, learners' skill can convert from controlled processing to automisation, and in doing so a focus on form is reduced with more emphasis on language use (p. 195). This form of communicative practice could involve the use of a task. However,

the task would have to be designed in a way that could elicit pre-taught linguistic forms, which, as we previously mentioned in 3.2.1, can occur through focused tasks. The procedure of presenting linguistic forms and then practising them through tasks is known as "task supported language teaching (TSLT)" (Ellis 2003, p. 147) (see figure 2).

Figure 2. *Task-supported language teaching (TSLT) (Ellis, 2003, p. 147)*

Presentation of declarative knowledge → (controlled processing)	Communicative practice through focused tasks	→	Communicative language use (procedural knowledge/ /automatic processing)

This method is similar to PPP and other types of skill learning theories in that proceduralisation is achieved by initially pre-teaching language form and then allowing learners to practice the form in a communicative setting using focused tasks. TSLT differs from TBLT in that Willis' (1996) framework does not advocate pre-taught linguistic forms, as attention to form comes after the task cycle. Seedhouse (1999) argues that TBLT's lack of attention to form prior to task performance would result in learners using minimal language to complete tasks and create impoverished language use. As we will see in the next sub-section, this may be the case with many Japanese learners who have had a lack of exposure using English during their education, and as a result, they may lack the confidence or ability to perform tasks using their own linguistic resources without the aid of explicit language guidance.

3.2.5 Issues with TBLT in Japan

TBLT is of particular relevance to Japanese educational contexts because since 2003, the Ministry of Education, Culture, Sports, Science and Technology in Japan (MEXT) has called for improvements in Japanese university students' use of English. MEXT was dissatisfied with traditional methods of language instruction that focused heavily on grammar translation as well as using reading and writing syllabuses which were not effective means for improving L2 oral proficiency for university graduates. As a result, educational institutions in Japan began responding to MEXT's goals by developing

English curriculums that focused more on developing students' L2 oral skills and implementing the use of tasks, particularly at the university level (see for example, Thompson & Jones, 2013). Although there have been numerous studies which have shown how oral tasks can be used to improve Japanese learners' L2 speaking skills (for example, Robinson, 2001, Thompson & Millington, 2012), they have all been carried out in experimental settings and there has been a lack of studies that have reported successful implementation of TBL programs within classroom contexts in Japan.

A bleaker picture is reported by Brown & Kikuchi (2009) whose study showed that Japanese students entering university did not appear to have benefitted from the communicative directives issued by MEXT because English language courses in senior high schools were still generally focused upon university entrance examinations which did not assess communicative competence but rather atomistic grammar, reading and writing. Consequently, there appears to be a reluctance to teach oral skills when university entrance exams target other language skills. Sakui (2004, 2007) reports other obstacles that hinder teachers' ability to foster communication skills, specifically large class sizes which cause classroom management difficulties for teachers who are unable to effectively monitor student L2 interaction, as learners can easily revert to the L1. These issues have resulted in a lack of exposure towards English oral communication for Japanese learners entering university, many of which may subsequently feel unprepared to engage in communicative tasks. Given students' exposure to more traditional methods of language learning that emphasize grammar instruction, Japanese learners may feel more comfortable with a TSLT approach to using tasks which provides pre-instruction of language form prior to performance, rather than using a more pure TBLT approach that requires learners to perform tasks without receiving any language input.

3.2.6 Summary: TBLT

As we have seen, there are different methods for using oral tasks, all of which provide advantages and disadvantages depending on the context. In the case of Willis' (1996) TBLT framework, attention to language form *after* task performance appears to clash with skill development theory and more traditional language teaching methodologies such as PPP which claim that development is more effective through the pre-teaching of

linguistic forms followed by practice and use of language in a communicative context. Tasks can have a part to play within PPP as they can serve as the communicate activity during the final performance stage, referred to as TSLT however, this is not 'pure' TBLT due to the pre-teaching of form which Willis' (1996) framework does not permit. Nevertheless, we can see that in using meaning-based tasks, language learning can occur through a focus on form *after* task performance, as favoured by Willis, or *prior* to task performance, as favoured by Dekeyser (1998). It is the latter phrase, referred to as 'pre-task planning' which is the focus of the present study and the topic of the next section.

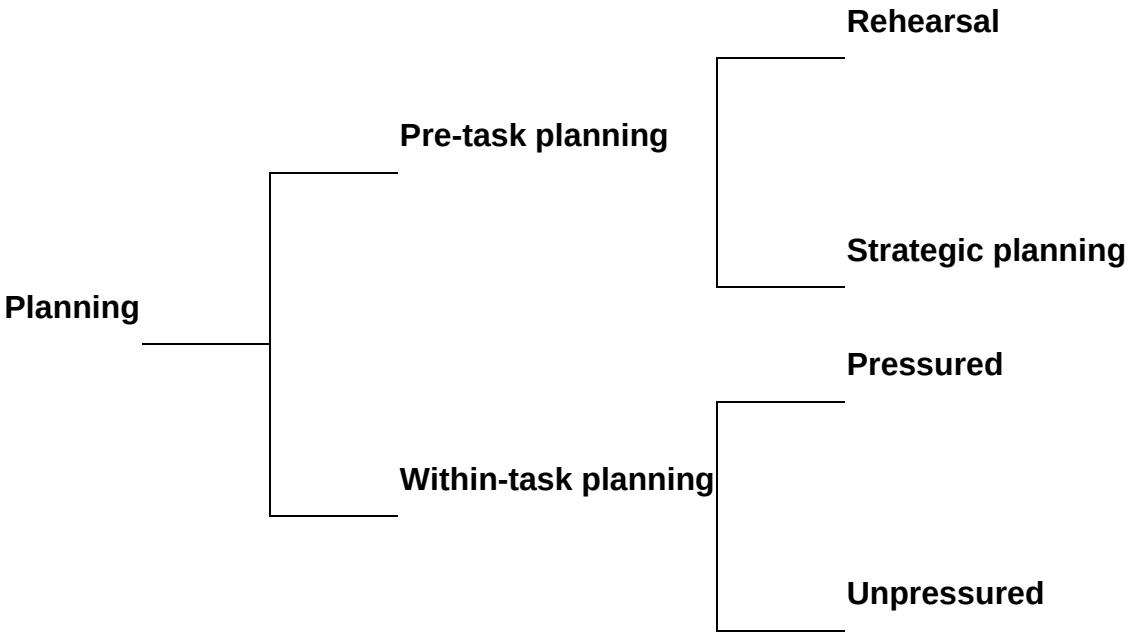
3.3 Pre-task planning

According to Ellis (2005a), task-based planning can be divided into the following categories:

- Pre-task planning
- Within-task planning

The distinction between the two forms is determined by when the planning occurs: prior to performing a task or during task performance. Pre-task planning takes place before the task, providing learners with time to prepare. Within-task planning (also referred to as 'online planning') relates to the time available for learners to prepare their speech *during* task performance. The length of time during within-task planning is dependent on whether the performance of the task is unpressured or pressured: the former implies that learners have time to prepare their speech during the task whilst the latter means that learners have limited time to plan whilst performing the task (see figure 3). However, for the purpose of this study we are focusing solely on pre-task planning.

Figure 3. *Task planning (Ellis, 2005, p. 4)*



"Pre-task planning is further divided into *rehearsal* and *strategic planning*. Rehearsal provides learners with an opportunity to perform the task before the 'main performance" (Ellis, 2005, p. 3). Rehearsal, otherwise known as 'task repetition' involves performing a task and then repeating it at a later stage. The initial performance of the task serves as a form of planning for the second performance, and as a result, the repeated performance is expected to yield gains in L2 production. "Strategic planning entails learners preparing to perform the task by considering the content they will need to encode and how to express this content". In other words, strategic planning provides time before a task where learners can plan what they want to say and how they want to say it.

The aim of this study is to investigate the effects of pre-task planning (strategic planning and task repetition) on L2 oral development in terms of fluency, accuracy and complexity. In order to do, we need to know how pre-task planning influences L2 speech, or more specifically, how these variables influence Kormos' (2011) bilingual model of speech production.

3.3.1 The implications of pre-task planning on Kormos' (2011) bilingual model of speech production

In order to consider the effects of pre-task planning on Kormos' (2011) bilingual model of speech production, we must also take into account Levelt's (1989) model of L1 production (which we earlier reviewed in 2.2) for two main reasons. First, both models are very similar in terms of the mechanisms involved in oral production i.e. conceptualisation, formulation and articulation, therefore comparisons between L1 and L2 production can be easily made. Second, as Kormos' (2011) model is only recently published, all previous planning studies have relied on Levelt's (1989) model as "a basis for considering what components of language production (spoken or written) learners focus on while planning and also for examining what effects planning strategies have on actual production" (Ellis 2005, p. 14). Consequently, reference to Levelt's (1989) model will allow us to compare the findings of previous planning research with the results of the present study. Second, the advantage of using Levelt's (1989) and Kormos' (2011) models is that they can enable us to formulate "relatively precise hypotheses about the effects that planning will have on task performance" (Ellis, 2005, p. 15). In the next sub-section, we examine two well-known psycholinguistic models within task planning research that attempt to predict the effects strategic planning can have on learners' attention and L2 oral performance.

3.3.2 The effects of strategic planning on learners' attention and L2 oral performance

SLA researchers have made predictions about how task conditions, such as strategic planning can influence learners' attention in different ways and how this impacts on task performance in terms of fluency, accuracy and complexity. Two of the most influential claims come from Skehan's (1998) Limited Capacity Hypothesis and Robinson's (2003) Cognition Hypothesis. In the case of the former, Skehan (1998) argues that learners have limited attentional resources, referred to as a 'single pool' which has adverse effects on fluency, accuracy and complexity when task demands are high, for example when there is no planning time. "The assumption is that more demanding tasks consume more attentional resources simply for task transaction, with

the result that less attention in available for focus on form" (p. 97). Skehan (1998) argues that complex tasks will result in learners' attending more towards meaning and the communicative aim of the task and less attention will be focused on language. Consequently, attention capacity limits forces the learner to prioritize one aspect of speech over another when performing complex tasks. As a result, tasks can result in gains in accuracy or complexity but not both. Taking Skehan's (1998) assumption that learners have a 'single pool' of attentional resources, Samuda & Bygate (2005) argue that it is possible to 'free-up' attentional resources by providing strategic planning "which reduces the processing load of subsequent on-line performance: speakers may have mentally organized the content; and/or worked on the formulation of aspects of the communication" (p. 39). In other words, strategic planning provides time for learners to attend to conceptualization (message content) and/or formulation (grammar encoding) which is then stored in memory and later produced during task performance as more fluent, complex and/or accurate L2 speech.

Robinson (2011) on the other hand, argues against limited attentional capacity processing and a trade-off between accuracy and complexity. He believes that learners can access 'multiple pools' of attention and that both aspects of speech can be improved by having learners perform more cognitively demanding tasks. For example:

> increasing the amount of reasoning a task requires, promotes greater effort at controlling production and more vigilant monitoring of output. This increased complexity leads to greater accuracy and complexity of L2 production when compared to performance on simpler task versions that require little or no reasoning. (p. 12)

According to Robinson (2011), a complex task which may involve learners explaining the reasons behind other people's actions, will increase the attention learners pay to their speech and their efforts at producing complex syntax, for example, cognitive state verbs - *he thinks that…she believes that*, compared to simple tasks that require no reasoning. In terms of L2 production, "complex task demands lead to greater effort at conceptualization and elicit the morphologically richer and structurally more complex syntactic mode" (Robinson, 2011, p. 14). Complex tasks, for example those that involve reasoning without planning time, will generate more elaborate communicative concepts at conceptualisation which in turn results in more complex *and* accurate L2 speech, at the cost of fluency. If however, planning time was permitted, positive effects on all three aspects of CAF would result.

We can see then that Skehan's (1998) and Robinson's (2003) theories provide contrasting views regarding the effects of strategic planning on oral task performance in terms of accuracy and complexity. Planning studies to date have lended weight to both models, as we will see in 3.3.5. Thus, in order to determine which of these two competing theories is more convincing, more research is needed.

3.3.3 Types of strategic planning: Guided vs Unguided

SLA researchers have investigated the way strategic planning can be manipulated in order to improve different aspects of L2 speech. One way is through guided planning which involves focusing learners' attention as they prepare for a task, for example, attending to specific aspects of grammar or vocabulary (Ellis, 2009). This type of instruction could be referred to as task-supported language teaching which was discussed in 3.2.4, as it favours pre-linguistic instruction. Guided planning could also focus on meaning by attending to the storyline or content of a task. Finally, it could involve attention to both language and meaning/content, as we will see in the following studies. Unguided planning, on the other hand, allows learners time to plan independently without any teacher-led assistance towards language or content. Thus, students are free to use their own linguistic resources to prepare for a task.

3.3.4 Guided and unguided planning studies

The question now is which type of planning (guided or unguided) is the most effective for promoting fluency, accuracy and complexity? During the past two decades, a significant amount of empirical studies have investigated the effects of strategic planning on L2 oral performance. In order to select appropriate studies for review, Norris and Ortega (2006) recommend that "research synthesis always includes an explicit articulation of how the relevant literature was searched and how primary studies were selected for review" (p. 6). In the case of this study, relevant sources were accessed from SLA literature including two references (Ellis, 2005, and Ellis, 2009a) which specifically reviewed all the key studies devoted to strategic planning over the

last twenty years. However, as this thesis is concerned with the effects of guided and unguided planning on L2 fluency, accuracy and complexity, only the studies which examined those variables were selected for review. For reasons of space, table 4 summarizes 12 of these studies followed by their results concerning CAF.

Table 4. *Guided and unguided planning studies*

Study	Context	Proficiency	Task Type	Strategic Planning Conditions	Language	Results (Speech Improvements)
Williams (1992)	Lesson with Korean St's speaking alone	Advanced	Poster presentation	Guided Instruction	Complexity, discourse markers	Complexity
Foster & Skehan (1996)	Non-class context with various EFL St's in dyads	Intermediate	Narrative, personal information, decision making	Guided & unguided	Fluency, accuracy, complexity	Fluency, Complexity, Accuracy
Mehnert (1998)	Non-class context with German St's in dyads	Intermediate	Instruction, exposition	Guided form and meaning	Fluency, accuracy, complexity	Fluency & Accuracy
Foster & Skekan (1999)	Non-class context with various EFL St's in groups	Intermediate	Decision Making	Guided form, content & Unguided planning	Fluency, accuracy, complexity	Fluency, Complexity Accuracy
Ortega (1999)	Non-class context with Spanish St's in dyads	Advanced	Narrative	Unguided planning	Fluency,accuracy, complexity	Fluency, accuracy, Complexity
Yuan & Ellis (2003)	Non-class context with Chinese St's in dyads	Intermediate	Narrative	Unguided planning & online planning	Fluency, accuracy, complexity	Fluency & Complexity
Sanguran (2005)	Non-class context with Thai St's speaking alone	Intermediate	Instruction, Argumentative	Guided form, meaning, & Unguided	Fluency, accuracy, complexity	Fluency, Complexity, Accuracy
Ortega (2005)	Non-class context with Spanish St's in dyads	Intermediate and Advanced	Narrative	Unguided planning	N/A	N/A
Kawauchi (2005)	Non-class context with Japanese St's in dyads	Lower & Higher Intermediate, Advanced	Narrative	Guided & Unguided Planning	Fluency, accuracy, Complexity	Fluency, Complexity, Accuracy
Skehan & Foster (2005)	Classroom	Intermediate	Decision making	Guided & Unguided planning	Fluency, accuracy, complexity	Fluency, Complexity, Accuracy
Gilabert (2007b)	Laboratory	Low Intermediate	Narratives	Unguided planning	Fluency, accuracy, complexity	Fluency, Complexity
Mochizuki & Ortega (2008)	Non-class context with Japanese St's in Dyads	Beginner	Narrative	Guided & unguided planning	Fluency, complexity, Accuracy (relative clauses)	Accuracy

3.3.5 Emerging patterns on fluency, accuracy and complexity

The above studies have shown that different types of strategic planning (guided and unguided) can improve fluency, accuracy and complexity. The studies showed that learners use planning time to attend to conceptualization, formulation and articulation which enables them to perform tasks with improved L2 speech. The results of these studies fall under two main categories. Williams (1992), Foster & Skehan (1996, 1999), Mehnert (1998), Kawauchi (2005), Yuan & Ellis (2003), Gilabert (2007b), Mochizuki & Ortega (2008) suggest that strategic planning can benefit some but not all aspects of L2 speech, namely fluency and complexity, depending on the task-type and planning condition. These results generally imply a 'trade-off' effect between complexity and accuracy. However, other studies such as Ortega (1999) and Sangarun (2005) show that strategic planning produces gains in all three aspects of speech. Table 5 provides a breakdown of these results in relation to fluency, complexity and accuracy.

Table 5. *Strategic planning effects on CAF from the above studies*

Strategic Planning effects for Fluency		
Studies	**Effect**	**Major Findings**
Williams (1992)	N/A	N/A
Foster & Skehan (1996)	Strong	Fewer pauses than unplanned
Mehnert (1998)	Strong	Fewer pauses than unplanned
Foster & Skehan (1999)	Strong	Greater fluency than no planned
Ortega (1999)	Strong	Fewer pauses than no planned
Yuan & Ellis (2003)	Strong	More syllables produced than no planning
Sangaran (2005)	Strong	Number of syllables produced more than unplanned
Kawauchi (2005)	Strong	Number of syllables produced more than unplanned
Gilabert (2007b)	Strong	Number of syllables produced more than unplanned
Mochizuki & Ortega (2008)	None	No change in the number of syllables produced
Strategic Planning effects for Complexity		
Studies	**Effect**	**Major Findings**
Williams (1992)	Strong	Number of clauses produced more than unplanned
Foster & Skehan (1996)	Strong	Number of clauses produced more than unplanned
Mehnert (1998)	None	No change in number of clauses produced
Foster & Skehan (1999)	Strong	Number of clauses produced more than no planned
Ortega (1999)	Strong	More words produced
Yuan & Ellis (2003)	Strong	More verb forms and clauses produced
Sangaran (2005)	Strong	Number of clauses produced more than unplanned
Kawauchi (2005)	Strong	More subordinate clauses produced
Gilabert (2007b)	Strong	Greater lexical richness than unplanned
Mochizuki & Ortega (2008)	Min	No change in the number of clauses produced for 2/3 measures
Strategic Planning effects for Accuracy		
Studies	**Effect**	**Major Findings**
Williams (1992)	None	No change in the number of error-free clauses produced
Foster & Skehan (1996)	Strong	More error-free clauses produced than unplanned only with one task type
Mehnert (1998)	Min	More error-free clauses produced than unplanned
Foster & Skehan (1999)	Strong	More error-free clauses produced under guided planning
Ortega (1999)	Strong	More error-free noun modifiers
Yuan & Ellis (2003)	None	No change in the number of error-free clauses produced
Sangaran (2005)	Strong	More error-free words produced than unplanned
Kawauchi (2005)	Min	More past tense makers produced with lower-level learners
Gilabert (2007b)	None	No change in number of self-repetitions
Mochizuki & Ortega (2008)	Strong	More error-free relative clauses produced under guided planning

These findings highlight a number of implications regarding the effects of strategic planning. First, it is evident that strategic planning is a worthwhile pedagogic tool for developing learners L2 oral skills as it produces clear gains in fluency and complexity, and on occasion accuracy. Second, the majority of the results lend weight to Skehan's (1998) Limited Capacity Hypothesis which claims that planning results in gains in fluency and complexity or fluency and accuracy and that a trade-off exists between accuracy and complexity. Thus, although we defined fluency, accuracy and complexity as distinct aspects of L2 speech, this does not mean they do not interact with each other. Furthermore, we must also take into account Housen et al.'s (2012) warning from 2.4 that discrepancies in CAF findings could also be due to a lack of clarity and consistency with the measures used in previous studies.

3.3.6 The impact of guided vs unguided planning on CAF

Mochizuki & Ortega (2008) note that "a particular concern in this strand of research has been with fostering linguistic accuracy, an area in which planning effects have resisted firm conclusions" (p.15). As we have seen so far, strategic planning appears to benefit fluency and complexity, however there are mixed results concerning accuracy. Consequently, guided planning has been manipulated to specifically target this aspect of L2 speech. To illustrate this, table 6 and 7 compare the findings of guided planning against unguided planning.

Table 6. *Guided Planning Results*

Guided Planning	Fluency	Complexity	Accuracy
Williams (1992)	N/A	Sig. Effect	No effect
Foster & Skehan (1996)	Sig. Effect	Sig. Effect	Sig. Effect
Mehnert (1998)	Sig. Effect	No effect	Sig. Effect
Foster & Skehan (1999)	Sig. Effect	Sig. Effect	Sig. Effect
Kawauchi (2005)	Sig Effect: all proficiency levels.	Sig. Effect	Sig. Effect: low intermediate only
Sangarun (2005)	Sig. Effect	Sig. Effect	Sig. Effect
Mochizuki & Ortega (2008)	No effect.	Sig Effect: with 1 syntactic measure only	Sig. effect

Generally, what we can see from table 6 is that guided planning facilitates significant gains in fluency, accuracy and complexity.

Table 7. *Unguided Planning Results*

Unguided Planning	Fluency	Complexity	Accuracy
Ortega (1999)	Sig. Effect	Sig. Effect	Sig. Effect
Yuan & Ellis (2003)	Sig. Effect	Sig. Effect	No effect
Gilabert (2007b)	Sig. Effect	Sig Effect: Lexical complexity only.	No effect
Mochizuki & Ortega (2008)	No Effect	No effect	No effect

Table 7 tells us that, on the whole, unguided planning produces significant effects in fluency and complexity but no effect for accuracy. Thus, the difference between the two types of planning conditions lies with accuracy in that guided planning appears to be more effective. Indeed, Mochizuki & Ortega (2008) comment that "in order to maximize the effectiveness of planning time some sort of guidance is beneficial, particularly when increased accuracy is the goal" (p. 15). As guided planning involves attention to language form and/or content, we can see that pedagogically, if we are

looking to improve all *three* aspects of learners' speech, attention to language form and/or content is crucial when allowing planning time. However, Mochizuki & Ortega (2008) point out that the challenge for teachers and SLA researchers is to decide on the type of guided planning that would benefit learners' proficiency. We now turn to examine Mochizuki & Ortega's (2008) study as it plays a crucial part in the methodology of this thesis.

Mochizuki & Ortega (2008) investigated the effects of strategic planning with guidance to a specific language form using beginner level students. The participants were 112 Japanese high school students who, in dyads, took part in a monologic narrative task. The task itself was designed specifically for the study in order to elicit a form known for its difficulty in oral production: relative clauses. Each speaker and listener were handed a set of pictures that described a story. Due to the students' low proficiency level, an audio recording of the story was played to the speakers in the L2 so as to aid their preparation in using relative clauses during the task.

The participants were split into three groups each with different planning conditions. One group received no planning time in which the speaker had to re-tell the story immediately after listening to the audio. The other group received five minutes of unguided planning time. The final group received five minutes of guided planning in which they received a handout containing written examples of simple relative clause types, for example "object-subject (OS, I like the dog which has long ears), object-direct object (OO, I want the dog which the little girl has in her arms), and subject-subject (SS, the dog which has long ears looks friendly" (Mochizuki & Ortega, 2008, p. 19), and instructions to try and use the grammar in the task. A focused task was designed to elicit seven obligatory contexts of the form within the story. Task performance was measured against fluency, complexity and accuracy. However, unlike previous planning studies, accuracy was measured by the quality of relative clauses produced using a six point rating scale.

The findings showed a low count in relative clause production amongst all the participants, "thirty-six percent of the speakers produced four or more of the expected seven relative clauses in their L2 narratives, whereas 27% produced between three and one relative clauses only, and 36% produced no relative clause at all" (Mochizuki & Ortega, 2008, p. 24). The guided planners produced significantly more relative clauses as well as more accurate use of the targeted forms compared to the unguided and no

planning groups. The study showed that the unguided planners did not have more success in producing the forms compared to the no planners.

In terms of fluency and complexity, the two planning conditions did not provide any significant gains compared with the control group, contrasting with previous studies on the benefits of strategic planning. In terms of complexity, the guided planning group produced significant gains in respect to one measure of complexity; 'relative clauses per t-unit' compared to the unguided planners and the control group. However, there were no significant improvements from both planning conditions in terms of the two other complexity measures used 'dependent clauses per t-unit' and 'words per t-unit'. These mixed results highlight the issues regarding complexity measures which we looked at in 2.4.6. Mochizuki & Ortega (2008) attributed the lack of gains in complexity and accuracy from the unguided planning conditions to the beginner proficiency level of the learners as they may not have had enough declarative knowledge to capitalize on the planning time available to be able to produce complex grammatical structures in the L2.

In conclusion, this study showed that guided planning which draws learners' attention towards a specific form can lead to improvements in accuracy but not fluency and complexity. However, as Ellis (2009a) points out, "the issue in such studies is what students actually do when they are asked to plan, but this has been rarely investigated (Ellis, 2009, p. 492). As this study did not investigate the strategies used by learners, it is unclear whether learners of low proficiency follow guided planning instructions as expected. Finally, the study is limited as it is unable to report the effects that guided planning may have on a repeated performance of the task, or on learners' development of relative clause use over time.

Ellis (2009a) points out that "there are some obvious limitations in the studies carried out to date. One of the most serious is the lack of information about what learners actually do while they plan" (p. 505). This is crucial because it cannot be assumed learners actually plan when given instructions. If they do plan, do they attend to meaning or form? The only study which has provided an in depth look at the strategies learners use during strategic planning is Ortega (2005), which we shall now examine.

3.3.7 Ortega's (2005) investigation into learners' planning strategies

Ortega's (2005) study analysed learners' meta-cognitive responses from post-task interviews from her previous planning studies (Ortega, 1995; 1999) (cited in Ortega, 2005) in order to find out learners' perceptions of strategic planning as well as the cognitive processes that they engage in during planning. Adult ESL learners of Spanish participated in these studies, the 1995 group consisted of 14 intermediate level learners and the 1999 group involved 32 advanced learners. In both studies, the students performed two monologic narrative tasks in dyads; one task with unguided planning time, the other task with no planning.

The post-task interviews revealed how individual learner differences play a major role in determining how learners perceive the benefits of planning as well as the strategies they use. For example, some learners had a natural inclination towards communication when speaking in the L2 whereas other learners were more concerned with speaking accurately and this influenced their opinions regarding the benefits of planning. Those students that had a strong orientation towards communication "seemed to accept error and error correction as inherent to their being non-native speakers of the language, and as part of a gradual process of second language learning" (Ortega, 2005, p. 92). These learners did not seem too concerned with accuracy, and furthermore, they did not see the benefits of planning in order to attend to form, but rather they preferred to speak without the assistance of strategic planning.

Other learners were more inclined towards accuracy as they "were anxious about making mistakes, and they seemed to view L2 learning as a prolonged effort to reach a hundred percent correctness" (Ortega, 2005, p. 93). As a result, these learners appreciated the opportunity to plan in an attempt to focus on form and practice their speech. In addition, as the tasks involved communicating to a listener, certain learners would "prioritize getting the message across to the listener over being accurate, fluent, or complex" (Ortega, 2005, p. 105). For example, some learners were reluctant to use complex grammar or vocabulary whilst others avoided correcting themselves during performance for fear of being misunderstood.

Ortega (2005) noted that although there were cases of learners attending to form or meaning during planning, there were also numerous instances of learners attending to "form-in-meaning" (p. 106), that is, certain learners "seemed to pay attention to the

inextricable relationship between form and meaning, simultaneously holding in long-term memory considerations regarding the message to be conveyed and the essential formal resources to convey it" (p. 106). In other words, strategic planning afforded the time for learners to weigh up the communicative task demands *and* attend to the language required to complete it. The study showed that certain learners were aware of the communicative nature of the task (a story-telling narrative) yet they also attended to form without being instructed to do so.

3.3.8 A critical review of Ortega's (2005) study

According to Samuda & Bygate (2008), Ortega's (2005) study "is unique in the current literature in exploring in depth the perceptions and beliefs about a task-based procedure which they have just experienced" (p. 174). Although the study was able to report learners' attention towards form and meaning, the unexpected finding of learners attending to form-in-meaning is supported by DeKeyser, Salaberry, Robinson, and Harrington who also argue that "simultaneous attention to form and content is clearly possible" (2002, p. 809). Long & Robinson (1998) describe how attention to form and meaning do not have to be mutually exclusive as "this is similar to what happens when native speakers who are good writers pause to consider the appropriate form of address to use when composing a letter" (p. 23). However, an important consideration regarding these viewpoints is proficiency. The participants in Ortega's (2005) study were upper-intermediate and advanced level learners, and consequently would have sufficient L2 knowledge to be able to simultaneously attend to form whilst processing language meaningfully. However, this would not be possible with lower level proficiency learners who would only be able to attend to form and meaning separately.

A limitation of Ortega's (2005) study is that it involved unguided planning only therefore it is unclear what learners would do if they were provided with guided planning. For example, do learners follow instructions as expected? Finally, despite the fruitful insights into the strategies learners use during planning, Ortega (2005) only examined learners' perceptions at a specific point in time. Consequently, we do not know whether learners' planning strategies change on subsequent performances or whether any patterns emerge as learners progress with more complex tasks. Knowing

this would provide valuable information in the way learners attend to different aspects of their speech over time.

3.3.9 Summary: strategic planning

We began this section by defining pre-tasking into its sub-components; strategic planning and task repetition and how different types of planning can allow learners time to focus on conceptualisation, formulation and articulation to improve L2 performance. We saw how psycholinguistic models attempt to predict how planning can improve different aspects of L2 speech. We then distinguished guided planning from unguided planning and saw how planning studies appear to facilitate gains in fluency and complexity whist gains in accuracy only appear possible if some sort of guidance is provided. However, these studies seem limited in that they only report the benefits of guided or unguided planning on immediate production use at specific points in time as there appears to be a lack of research that has investigated the effects of guided and unguided planning on CAF over time. The section concluded by reviewing the strategies learners use during planning which showed that learners appear to plan according to their L2 oral orientations towards accuracy and communication. However, there appears to be no research that has examined learners' planning strategies on sequenced tasks therefore we don't know how learners orient themselves towards planning over time. The following section will now discuss the other component of pre-task planning: task repetition.

3.4 Task repetition and task sequencing

The purpose of this section is to discuss how different forms of task repetition can benefit CAF. We first begin by defining what we mean by task repetition. The following sub-section then reviews Bygate's (2001b) study that investigated the construct. 3.4.2 then introduces a different form of task repetition referred to as task complexity which involves sequencing tasks according to an increase in their cognitive demands. In 3.4.3 we discuss the components of task complexity that are relevant for the present study. Finally, 3.4.4 reviews L2 developmental studies involving task complexity.

3.4.1 Task Repetition

Task repetition can be defined as "repetitions of the same or slightly altered tasks – whether whole tasks or parts of a task." (Bygate & Samuda, 2005, p. 43). The past decade has seen the emergence of task repetition research, although not to the same extent as strategic planning. Key studies such as Bygate (2001b) and Bygate & Samuda (2005) show that "the repeated use of the same and similar communication tasks can affect processing in such a way as to be capable of fostering language development" (Bygate, 2001b, p. 29). However, for reasons of space, only one study can be reviewed. Bygate (2001b) carried out a study that reported the effects of task repetition on fluency, accuracy and complexity. Students were split into three groups: an interview group, a narrative group and a control group. In week one, all the students performed one narrative and one interview task. The narrative group would then perform two similar narrative tasks every two weeks over a 10 week period. The interview group would do the same with interview tasks. On week 10, all the students performed two narratives and two interview tasks, one of each was a repetition of the task they did in week one. The control group did not perform any tasks during the two week intervals. The results showed that "there is a strong effect for task repetition" (Bygate, 2001b, p. 42). Students were able to perform a repeated task with greater fluency and complexity than their initial attempt 10 weeks earlier. Despite gains in fluency and complexity, the study is limited in that the benefits of task repetition could not be transferred over to another task, in other words, the narrative group did not show gains in the interview task and the interview group did not produce gains in the narrative task. The failure to show how repetition can lead to gains with different task-types prompted Ellis (2009a) to suggest that this "should serve as a warning call to all planning studies" (p. 505).

Another form of task repetition that has emerged in SLA literature in recent years is task complexity. The distinction between the two is based on the concept of complexity which the next sub-section will discuss.

3.4.2 Task complexity and the Cognition Hypothesis

Task complexity involves sequencing the same type of task but increasing its complexity on subsequent versions. Robinson (2003) developed a theoretical framework for task complexity referred to as the Cognition Hypothesis. According to the claims of the Cognition Hypothesis, "distributing optimal task-based language use and learning opportunities over time, i.e. *task sequencing*, is done by designing and having learners perform tasks simple on all the relevant parameters of task demands first, and then gradually increasing their cognitive complexity on subsequent versions" (Robinson, 2010, p. 242). In other words, L2 development can be achieved by having learners move from simple to complex tasks along cognitive dimensions. For example, moving from 'here and now' tasks which involve using the present tense to 'there and then' tasks which involve using the past tense requires the learner to differentiate between temporal states of reference in the L2 (present tense versus past tense) and to use deictic terms accordingly (*here, there, this, that*). Robinson (2005) notes this sequence of moving from present to past occurs in L1 acquisition of English where children acquire the use of the present tense before the past tense, as well as in L2 acquisition where past tense morphemes tend to be acquired later than the present tense. Consequently, sequencing tasks according to increasing cognitive complexity reflects the development observed in L1 and L2 acquisition (Kormos, 2011). Robinson (2001) argues that the process of sequencing tasks according to an increase in complexity acts as "a more powerful influence on production than repetition of task versions" (p. 40). As Robinson's (2003) proposal attempts to 'push' learners output by performing more complex tasks, as opposed to simply repeating the same task, it appears a more optimal strategy for promoting L2 development in terms of fluency, accuracy and complexity.

Robinson & Gilabert (2007) provide a framework for designing tasks which includes a full taxonomy of task features, criteria and procedures that can be used when attempting L2 developmental studies (see appendix A). This framework consists of three main categories, however, for the purpose of this study, we shall examine one aspect of the model, which is essentially the sole factor within the framework for sequencing tasks: task complexity.

3.4.3 Task complexity

Task complexity relates to the cognitive demands of tasks, of which there are two main influences: "*resource-directing* dimensions, and *resource-dispersing* dimensions" (Robinson, 2005, p. 4). The former relates to the content or linguistic demands that tasks place on learners which can be manipulated by changing the variables associated with it. For example, tasks may involve intentional reasoning demands, that is, they require the learner to explain the actions or thoughts of other people in a story. In L2 English, this can involve the use of psychological cognitive state verbs such as *he thinks..., she believes* etc as well as the additional L2 structures that accompany them, for example, relative clauses; *he thinks that he likes the dog which has long hair*. Robinson (2010) argues that increasing tasks along *resource-directing* dimensions can increase the attention learners pay to their speech and their efforts at producing complex syntactic language.

Robinson (2007) illustrated this by carrying out a study using three interactive tasks that were designed to increase in complexity along intentional reasoning demands. 42 Japanese learners participated in the study, and in dyads, they were required to perform one simple, medium and complex task. The first task involved narrating a story about a person's intention to build a house. The subsequent tasks increased in intentional reasoning demands by containing more instances where the character responded to other people's opinions. The study used specific measures for all psychological state terms that were expected to be used, for example *he thinks.., he believes...* The results showed that cognitive state terms were produced more frequently using ratio measures of 'cognitive state terms per clause' in the complex task compared to the other two versions, resulting in more complex speech. Robinson (2005) argues that "increasing complexity along these dimensions therefore has the potential to direct learners' attentional and memory resources to the way the L2 structures and codes concepts, so leading to interlanguage *development*" (p. 4). In other words, sequencing tasks that increase along resource-directing dimensions, for example, increasing intentional reasoning demands, directs learners' attention to aspects of language that are required to meet the linguistic demands of complex tasks and results in more complex and accurate output. As we will see in the following chapter, cognitive state verbs and

intentional reasoning speech will play an important role as the target language for learners to produce in our pilot study.

Kormos (2011) supports Robinson's (2005) claims concerning task complexity and L2 development in relation to her psycholinguistic model of L2 speech production which we examined in 2.3.1. "Tasks increasing in cognitive complexity require L2 learners to activate complex concepts and more detailed schema of the communicative functions….and hence drive learners to express more complex relations among the activated concepts" (Kormos, 2011, p. 53). If tasks that increase in resource-directing dimensions active more complex concepts, how does this process influence L2 speech production? Kormos (2011) notes that "the complexity of tasks has a large influence on the macro-planning stage, where concepts are selected and relations among them are encoded". We saw earlier in 2.3.1 that the learner generates concepts during conceptualization in the form of a communicative pre-verbal plan, and these concepts activate associated lemmas from the learner's mental lexicon for syntactic encoding during formulation. Consequently, elaborate concepts will generate more sophisticated vocabulary and grammatical structures resulting in higher levels of complex and accurate language. Thus Kormos (2011) claims "tasks that are complex on resource-directing dimensions induce greater lexical variety and higher syntactic complexity".

The other cognitive factor of task complexity: *resource-dispersing* dimensions concerns the performance demands that tasks place on learners which can be manipulated by altering the variables associated with it such as strategic planning. For example, sequencing tasks where planning time is reduced serves to promote "greater *control* over, and faster access to existing interlanguage systems of knowledge" (Robinson, 2010, p. 248). This process 'pushes' the learner to access and retrieve linguistic resources at a faster rate in order to produce L2 output without time delays. Increasing dispersing complexity therefore enhances automatisation of what learners already know and primes learners to perform tasks more fluently under the normal time pressures of everyday speech. As we know from 3.2.4 which examined Anderson's (2000) skill theory, this process is referred to as proceduralisation in which declarative knowledge is transformed into procedural knowledge, enabling the learner to use their linguistic resources at a faster rate resulting in a more fluent performance. However, Robinson (2010) points out that increasing complexity along resource-dispersing dimensions however, does *not* promote interlanguage development (as opposed to

resource-directing dimensions) but rather, it enables the ‘speeding up’ of a learner’s interlanguage processing system, benefitting fluency.

Together these two dimensions (resource-directing and resource-dispersing) are the sole factors for sequencing tasks in order to develop complexity, accuracy, and fluency and to prepare learners for real-world task performances.

3.4.4 The effects of task complexity on L2 development

According to Baralt (2010), there appears to have been seven studies that have investigated the effects of task complexity on L2 development, all involving increases in complexity along resource-directing dimensions. These studies are briefly illustrated in table 8 which indicates the task complexity variable used for development, the targeted language form, the task-type and brief results.

Table 8. *Task Complexity L2 developmental studies (Baralt, 2010, p. 87)*

Study	Task Complexity variable	Dyad Type	Task	Results
Revesz & Han (2006)	(+-) then and there	ESL learners and researcher	Describing a video	Gains in the English past progressive
Nuevo (2006)	(+-) causal reasoning	ESL learners	Narrative & Decision making task	No gains in English past tense
Revesz (2009)	(+-) here and now	EFL learners and researcher	Description task	Gains in English past progressive
Kim (2009)	(+-) causal reasoning	EFL learners	4 interactive tasks	Gains in question formation and past tense
Kim & Tracy- Ventura, in press)	(+-) causal reasoning	EFL learners	4 different tasks	Gains in question formation and past tense
Revesz, Sachs & Mackey, in press)	(+-) here and now	EFL learners and researcher	Description tasks	Gains in past progressive tense
Baralt (2010, unpublished doctoral dissertation)	(+-) intentional reasoning	ESL learners	A Narrative	Gains in Spanish past subjunctive

Baralt (2010) points out that the above studies investigated L2 development through either recasts or through 'language related episodes' in which learners question their language use during interaction and are provided with correct feedback from an interlocutor. As a result, these studies used two-way interactive tasks in which L2 learning took place either *during* or *after* task performance. Thus, we do not know the

effects that strategic planning could have *prior* to task complexity on L2 development using one-way monologic tasks. As a result, the present study has chosen to examine the effects of strategic planning and task complexity on L2 oral development.

3.4.5 Summary: task repetition and task complexity

This section began by discussing research related to task repetition and how repeating tasks benefits L2 performance in terms of CAF. We then moved onto discuss another form of task repetition, referred to as task complexity which, according to the claims of Robinson's (2003) Cognition Hypothesis, involves sequencing tasks so they increase in cognitive complexity. Task complexity involves two dimensions: resource-directing and resource-dispersing. The former relates to the linguistic demands of a task, for example, intentional reasoning demands. Increasing the complexity of tasks along resource-directing dimensions directs learners' attention to the language code required to complete a task and promotes opportunities for more complex and accurate language. On the other hand, resource-dispersing dimensions impacts on performance, for example, whether or not planning time is allowed. Increasing the complexity of tasks along resource-dispersing dimensions (reducing planning time) primes learners to perform tasks under real world conditions by producing output at a faster rate, thus benefitting fluency. Consequently, Robinson (2010) argues that increasing the complexity of tasks along resource-directing and resource-dispersing dimensions can provide optimal levels for L2 oral development. Although as we have seen, L2 developmental studies involving task complexity appear limited in that no research has appeared to investigate the effects of strategic planning with task complexity using monologic narratives over time.

3.5 The effects of strategic planning and task complexity on L2 oral development

The purpose of this section is to examine how strategic planning and task complexity could facilitate L2 developmental gains in CAF. The following sub-section specifically addresses this matter, 3.5.2 then looks at the type of knowledge gained from a study that investigates the effects of planning on L2 oral development in terms of CAF. Finally, from 3.5.3 – 3.5.5

we examine the main linguistic form chosen for the present study as an indicator of L2 oral development: English relative clauses.

3.5.1 Strategic planning, task complexity and developments in CAF

As mentioned in 2.4.7, developmental changes in a learner's internal L2 system can be recognized by improvements in fluency, accuracy and complexity (Housen et al., 2012). For example:

- Improvements in fluency indicate proceduralisation of L2 knowledge as the learner can access L2 resources with reduced time delays, thus showing greater control of their L2 system.
- Improvements in complexity can be associated with developments in L2 knowledge as a result of learning new grammatical structures.
- Improvements in accuracy represent restructuring of a learner's internal L2 system to meet target-like performance and eliminate errors during production.

Ellis (2009a) however, chooses to dichotomise fluency development with that of complexity and accuracy as he argues that complexity and accuracy can lead to acquisition of linguistic knowledge whereas fluency development does not. Ellis (2009a, p. 504) illustrates this by outlining three types of acquisition:

(i) the acquisition of new linguistic features
(ii) the restructuring of existing L2 knowledge
(iii) the development of greater control (accuracy) over existing linguistic features

Ellis (2009a) believes that strategic planning has little influence on (i) especially with monologic tasks as they involve no interaction and opportunity for negotiation of meaning that could facilitate the acquisition of new linguistic features. Strategic planning could however, facilitate acquisition points (ii) restructuring, and (iii) improvements in the control of language (accuracy). Skehan (1998) argues that strategic planning facilitates the restructuring of existing L2 knowledge due to the positive consequences it has on L2 complexity. For example, we know from 3.3.2 that strategic

planning provides time for learners to conceptualize and formulate more complex utterances during task performance. This has led Ellis (2009a) to conclude that "more complex production will lead to acquisition" (p. 504). In terms of fluency development, both Ellis (2009a) and Skehan (1998) argue that it is less dependent on the proceduralisation of L2 rules as Housen et al. (2012); Kormos (2011); Robinson (2010) claim, but instead fluency is more dependent on access to formulaic chunks that require minimal processing and enable effective online communication to take place. As a result, Ellis (2008) argues that "the development of fluency can take place independently of the acquisition of linguistic resources" (p. 504). Strategic planning would benefit fluency development by strengthening the links between formulaic chunks so learners can gain faster access to them during formulation and task performance.

In terms of L2 oral development then, the key issue between SLA researchers appears to be the relationship between fluency and the proceduralisation of L2 knowledge. As we have seen in 2.4.7 with Housen et al. (2012), and in 3.4.3 with Robinson (2010) and Kormos (2011), fluency develops through automatisation of L2 rules and formulaic chunks whereas Ellis (2009a) and Skehan (1998) believe fluency is more dependent on access to, and accumulation of, formulaic chunks. Nevertheless, strategic planning would appear to benefit fluency, accuracy and complexity over time.

So how can we test L2 oral development? In the case of task planning, development can be measured by having learners perform a pre-test followed by treatment that involves planning and then an immediate and a delayed post-test (Ellis, 2005). Comparisons can be drawn with a control group of learners who take part in the tests but not the treatment. The only planning study to carry out this design was Bygate's (2001b) task repetition study which we reviewed in 3.4.1 and it involved repeating oral tasks every two weeks over a ten week period. However, a limitation of this study was that it used general measures for fluency, accuracy and complexity which could report general linguistic change but not the development of specific linguistic features. Ellis (2005a) points out that in order to analyze the effects of planning on specific forms, it is necessary to use focused tasks designed to elicit their use, as in the narrative used in Mochizuki & Ortega (2008) for eliciting relative clauses. In order to develop learners' use of a linguistic form over time, focused tasks would need to be sequenced together in a principled way, for example, by following the claims of Robinson's (2003) Cognition Hypothesis which, as discussed in 3.4.2, involve

sequencing tasks that increase in complexity. Comparisons would then be made between the pre- and post-tests for gains in CAF. However, as mentioned with our definition of fluency in 2.4.1, real-time speech is generally spontaneous therefore planning time would need to be prohibited during the pre- and post-tests to reflect natural speaking conditions. Consequently, if a learner could produce a targeted linguistic form during a task that did not involve planning time, what sort of L2 knowledge would the learner have acquired?

3.5.2 L2 knowledge representation

The ability to produce grammatical structures during communication without thinking requires the acquisition of implicit knowledge. According to Ellis (2008, p. 418), "implicit knowledge is intuitive, procedural, systematically variable, automatic, and thus available for use in fluent, unplanned language use". Thus, implicit knowledge of a targeted form would seem to be acquired if a learner could produce the form during a post-test task that did not involve planning. Implicit knowledge contrasts with 'explicit knowledge' which Ellis (2008) defines as "conscious, declarative, anomalous, and inconsistent (i.e. it takes the form of fuzzy rules inconsistently applied) and generally only accessible through controlled processing in planned language use" (p. 418). Thus, explicit knowledge appears to resemble the knowledge learners have when engaged in strategic planning where they can consciously attend to language prior to task performance. The characteristics of implicit and explicit knowledge therefore seem to resemble procedural and declarative knowledge. As we discussed in 3.2.4, procedural knowledge enables the learner to perform a skill automatically (as would be the case with implicit knowledge) whereas declarative knowledge requires more conscious control during performance and is only factual based (as with explicit knowledge). However, an important point to consider when distinguishing implicit and explicit knowledge is Dekeyser's (2003) argument that explicit knowledge may be proceduralised to the extent that it could be identical to implicit knowledge, in that both forms of knowledge could be accessed during unplanned conditions. Thus it could be argued then, that a learner may have proceduralised a targeted form to the point where he/she could access the form quickly from their explicit knowledge store during an unplanned test.

Ellis (2008) argues that "the terms 'explicit/implicit' label the type of knowledge learners possess according to whether it is conscious or intuitive, whereas the terms declarative/procedural concern the degree of control the learner has over both types of knowledge" (p. 430). Although Ellis (2008) does point out that this is not a clear cut issue and SLA researchers differ in their interpretations of implicit and explicit knowledge. Eysenck (2001) (cited in Ellis, 2008, p. 408) suggests they can be considered the same:

> It remains the case that declarative memory resembles explicit memory, in that it involves the integration or linkage of information. In contrast, procedural memory still resembles implicit memory, in that it involves specific forms of processing. (p. 213)

For the sake of clarity, we shall follow Eysenck's (2001) view that declarative and explicit knowledge represents factual L2 knowledge whereas procedural and implicit knowledge represent automatic forms of language processing. Let us now turn to the linguistic form which the present study will use as an indicator of accuracy and complexity development: relative clauses.

3.5.3 English relative clauses

A considerable amount of research has been carried out on English relative clauses over the past fourty years, mainly due to the fact that relative clauses (RCs) are universally linguistic as many of the world's languages contain the grammatical feature (Ellis, 2008). SLA researchers have therefore been interested in comparing RCs from different languages "in order to identify what features and structures they have in common" (Ellis, 2008, p. 557). Based on the data of fifty languages containing RCs, Keenan and Comrie (1977) formulated the Noun Phrase Accessibility Hierarchy Hypothesis (NPAH) which outlines the universal order difficulty for RC acquisition across all L2 languages that contain the feature. According to the NPAH, RCs which contain a relative pronoun that operates as the subject of a clause are considered to be the easiest to acquire in the L2 (also referred to as 'most accessible'). When the relative pronoun operates as the object of a clause, or acts a possessive, the RC sentence type becomes more difficult to acquire. The NPAH consists of six relativized functions in a hierarchical order of difficulty with the 'object of a comparison' considered to be the most difficult RC type (see figure 4):

Figure 4. *Relative clause order of difficulty* (*Keenan & Comrie, 1977, p. 66)*
SU > DO > IO > OBL > GEN > OCOMP

'SU' refers to subject, 'DO' refers to direct object, 'IO' refers to indirect object, 'OBL' refers to oblique object (otherwise known as object of a preposition (OPREP)), 'GEN' refers to genitive or possessive (POSS), and 'OCOMP' refers to object of a comparison. Finally, the symbol '>' represents more accessible than. Izumi (2003) provides example sentences of each RC type in which the relative pronoun is located in both the subject and object positions (see figure 5):

Figure 5. *Example sentences of each RC type (Izumi, 2003, p. 288)*

Subject	S	The woman who speaks Russian fluently is my aunt.
	DO	The car which the man drove is very fast
	IO	The man who(m) I gave the book to is my colleague.
	OPREP	The woman who(m) Bill is looking for is beautiful.
	GEN	The man whose car broke down is my boss.
	OCOMP	The mountain which Mt.Fuji is higher than is Mt. Takao.
Object	S	The teacher liked the girl who passed the exam easily.
	DO	We like the coat which Mary wears.
	IO	Mary likes the man who(m) Tom wants to live with.
	OPREP	She is the woman who(m) Tom wants to live with.
	GEN	I know the woman whose husband is a professor.
	OCOMP	I know the hotel which Hilton is cheaper than.

Although other RC hypotheses have since been formulated, for example, Hamilton's (1994) SO Hierarchy Hypothesis (SOHH), and Kuno's (1974) Perceptual Difficulty Hypothesis (PDH), Ellis (2008) notes that the majority of EFL and ESL RC acquisition studies largely support the claims of the NPAH (for example, Izumi, 2003). We now turn to discuss the issues Japanese learners have with learning English RCs.

3.5.4 English relative clauses involving Japanese learners

Japanese learners are taught English RCs from the second year of junior high school however studies such as Mochizuki & Ortega (2008) which we discussed in 3.3.6 have reported difficulties in relation to Japanese learners' oral production of the form. Although English RCs are well known for their difficulty as an aspect of English grammar for all L2 learners (Swan & Smith, 2001), in terms of Japanese learners, this is compounded by the differences that exist between English RCs and Japanese RCs. Schachter (1974) highlighted three areas of difference that exist between the forms:

1. The location of the RC in relation to the head noun. Japanese RCs are situated to the left of the head noun phrase whereas English RCs are situated to the right. An example of a Japanese RC situated before the noun is provided by Ellis (2008, p. 563):
   Gakussi ga katta hon
   Student NOM bought book
   the book that the student bought…
2. English RCs use a subordinate marker *'that'* between the head noun phrase and the RC, or relative pronouns *'who', 'whom', 'which', 'whose'.* Japanese RCs do not consist of subordination markers or relative pronouns but instead involve subordinate affixes.
3. Japanese RCs consist of pronominal reflexes whereas English RCs do not. According to Schachter (1974, p. 209), if English RCs did contain these pronouns, they would resemble the following:
   Subject: the boy that *he* came
   Direct Object: the boy that John hit *him*
   Indirect Object: the boy that I sent a letter to *him*
   Object of Preposition: the boy that I sat near *him*

Schachter's (1974) study also showed how these differences seem to result in Japanese learners' avoidance of English RCs. The study involved an error analysis of 50 compositions of English RCs from Japanese, Chinese, Persian and Arabian learners who were of intermediate and advanced proficiency. The results of the study showed that

Japanese and Chinese learners produced significantly less RCs compared to the Persian and Arabian learners. Schachter (1974) attributed Japanese learners' avoidance of RCs to the differences that exist with RCs in their own L1. Persian and Arabian RCs on the other hand, have more similarities with English RCs, for example, they both occur to the right of the head noun phrase. This led Schachter (1974) to conclude that "the learner apparently constructs hypotheses about the target language based on knowledge he already has about his own language….if they are radically different, he will either reject the new construction or use it only with extreme caution" (p. 212).

Japanese learners' avoidance of RC production reported in Schachter (1974) appear to be supported by the findings of Mochizuki & Ortega (2008) which seems to be the only study that examined learners' production of RCs in natural language use during a task. The results were disappointing in the sense that the sample as a whole produced a low mean average of RCs despite the narrative's design to elicit the form. Mochizuki & Ortega (2008) attributed this to the lack of task essentialness for producing the form as the learners could complete the task by avoiding using RCs, as well as the fact that their beginner proficiency level may not have been high enough to benefit from the planning guidance towards the form.

3.5.5 English RC instruction

We know that a considerable amount of research has been carried out on RC acquisition so for the purpose of this study, we shall examine the most relevant ones in terms of how RC instruction leads to acquisition. Erkman et al. (1988) carried out a pre- post-test design to see whether instruction towards one particular RC type (either 'subject', 'object' or 'object of a preposition') would transfer across to improvements in the use of other RC types. Erkman et al. (1988) referred to the RC types that are more difficult to acquire, such as object of a preposition, as being more typologically marked than simple RC types such as subject and object RCs. Markedness relates to the similarities or differences that exist between the grammatical structures in the L2 and its equivalent in the L1. The less marked an L2 grammatical structure is, the more similar it is to its L1 equivalent and is subsequently considered to be easier to acquire. In the case of the RC order of difficulty according to the NPAH, Erkman et al. (1988, p. 5) illustrate the degree of markedness as follows:

Figure 6. *Degree of markedness*

Accessibility Hierarchy	
Subject	least marked
Direct Object	
Indirect Object	
Object of a preposition	
Possessive	
Object of a comparative	most marked

The study involved 36 ESL intermediate learners of mixed L1 background and the results showed that the group which received instruction to the more marked RC type: 'object of a preposition' performed the best across all three RC types in terms of number of errors made compared to the 'subject' group and the 'object' group. This led Erkman et al. (1988) to conclude that "maximal generalization of learning will result from the acquisition of relatively more marked structures" (p. 12). In other words, RC instruction towards more marked RC types, in this case 'OPREP' will result in greater learning effects across other RC types 'S' and 'O' compared to instruction towards simple RC types. These results were supported by Doughty (1991) which also reported the benefits of OPREP instruction for providing generalized learning effects across less marked RC types (subject and object).

Thus, the findings of Erkman et al. (1988) and Doughty (1991) appear significant if we are to investigate the benefits of guided planning and task complexity using relative clauses as a targeted form. As instruction towards more marked RC types such as OPREP appears to benefit the learning of more simple RC types such as S and O it appears pedagogically worthwhile to include this form of RC instruction into the present study.

3.5.6 Summary: L2 development

We began this section by discussing the effect planning has on the development of CAF. Whilst there is general agreement in SLA literature that planning can lead to acquisition through improvements in accuracy and complexity, there are mixed opinions regarding fluency development as some researchers argue fluency develops through the automatisation of linguistic rules, whereas others argue that it is more dependent on the use of formulaic chunks. We then saw how L2 development could be confirmed by testing implicit knowledge that would involve learners performing pre- post-tests that did not involve planning time. We then moved onto the linguistic form to be investigated in the present study: relative clauses. We described the differences that exist between English RCs and Japanese RCs which explain why Japanese learners are known for avoiding the form during oral production. Finally, we reviewed how instruction towards more marked RC types such as OPREP appear to have general learning affects over more simple RC types such as S and O. Mochizuki & Ortega (2008) seems to be the only study that has examined RC production during natural language use i.e. through the use of tasks without controlled tests. However, as that study examined beginner learners' use of simple RC types at a specific point in time, it remains to be seen whether guided planning can result in the accuracy *development* of simple *and* complex RC types for intermediate learners over time. Furthermore, an additional linguistic challenge for intermediate learners would be to include L2 structures that compliment relative clauses such as cognitive state verbs that were discussed in 3.4.3, for example '*he thinks* that he likes the dog *which has long hair'*. In doing so, we could see how effective guided planning could be in developing intermediate learners' use of complex syntactic language.

3.6 Chapter conclusion

The purpose of this chapter was to provide a backdrop into the research surrounding guided planning, task complexity and its effects on fluency, accuracy complexity. We have seen how there are different methodologies for using oral tasks, for example Willis' (1996) TBLT framework that focuses on form after task performance whereas as TSLT draws learners' attention to form prior to performance. We then looked at

strategic planning research which investigates the effects of guided and unguided planning on learners' performance of tasks in terms of fluency, accuracy and complexity. These studies appear to suggest that gains in accuracy in particular are only possible under guided planning conditions. However, as these studies only reported the immediate effects of strategic planning, there appears to be a gap in the literature regarding the effects of strategic planning on CAF with tasks that are sequenced over time. Furthermore, only Ortega (2005) was able to report the strategies learners use when planning, but as this study was carried out a specific point in time, there is a need for research to examine whether learners' strategies change as they prepare for more complex tasks over time as it is unclear whether "learners attend serially to complexity, accuracy and fluency through the cycles of repetition (or quasi-repetition) of a given task" (Samuda & Bygate, 2008, p. 249). Although we discussed the benefits of task repetition and the claims of Robinson's (2003) Cognition Hypothesis which provides a theoretical guideline for sequencing tasks through increased complexity, it appears no research has yet to examine the combined effects of strategic planning and task complexity on CAF over time. Furthermore, it remains to be seen whether specific linguistic forms, such as simple and complex RC types as well as the accompanying use of cognitive state verbs can be developed over time through the treatment of guided planning and task complexity, and tested under the conditions of natural, unplanned language use. The value of addressing these gaps in the literature will us allow us to see how learners'attention can be drawn towards linguistic forms known for their difficulty in oral production and then developed through task sequencing with potential gains in fluency, accuracy and complexity. Furthermore, we can discover how learners' orient themselves towards performing more cognitively demanding tasks which will help our understanding in attempting to design syllabuses that maximize Japanese learners' L2 oral development through the use of tasks. Consequently, the next chapter reports on a pilot study that addressed the following research questions:

1. To what extent does guided planning and task complexity facilitate L2 oral development in terms of accuracy involving English relative clause types OS and OO aswell as cognitive state verbs, fluency and syntactic complexity of Japanese intermediate and upper-intermediate learners of English?

Hypothesis one: Guided planning and task complexity will facilitate L2 development to a greater extent than unguided planning and task complexity, and a control group, in terms of accurate use of English relative clause types OS and OO.

Hypothesis two: Guided planning and task complexity will facilitate L2 development to a greater extent than unguided planning and task complexity, and a control group, in terms of accurate use of cognitive state verbs.

Hypothesis three: Guided planning and task complexity will facilitate L2 development to a greater extent than unguided planning and task complexity, and a control group, in terms of syntactic complexity.

Hypothesis four: Guided planning and task complexity will facilitate L2 development to a lesser extent than unguided planning and task complexity, and a control group, in terms of fluency.

2. What strategies do Japanese intermediate and upper-intermediate learners of English use when planning for oral narratives that increase in complexity over time?

Hypothesis five: The guided planners would initially focus on form, then as tasks increase in complexity they would gradually attend more towards meaning over time.

Hypothesis six: The unguided planners would initially focus on meaning, then as tasks increase in complexity they would gradually to attend more towards form over time.

4. THE PILOT STUDY

4.1 Introduction

This chapter reports on the research methods used to carry out a pilot study which was intended as a small scale replication of the main study in order to find out preliminary answers to the research questions posed at the end of the last chapter. In addition, the pilot study would provide the researcher with valuable experience needed to successfully manage a task-based study over time that involved task design, monitoring and recording students' performance of simple and complex oral tasks, carrying out post-task interviews, analysing quantitative and qualitative data, and then determining whether our hypotheses were confirmed or unconfirmed. According to Robson (2002) a pilot study "helps to throw up some of the inevitable problems of converting your design into reality" (p. 383). Thus the results of the study would enable us to see what amendments were needed for the main study in terms of tasks, participants, procedures, planning conditions and other variables. This chapter therefore serves as a pre-recquisite to the following chapter which is devoted to the methodology of the main study. As a result, many of the sub-sections in this chapter and the next are similar in that they contain descriptions of components that were used in both the pilot study and the main study, for example, treatment tasks. In order to avoid repetition, this chapter briefly outlines and justifies the materials and measures used in the pilot study whilst the following chapter discusses them in more detail.

We begin in 4.2 by looking at the methodology required to answer our two research questions which do not change to a great extent in the main study. As research question one investigates learners' L2 oral performance in terms of fluency, accuracy and complexity, we discuss the benefits of quantitative research. As research question two examines learners' planning strategies, we also look at the benefits of qualitative research. As the overall study involves analysing learners' performance over time, we also discuss the merits of longitudinal research in 4.2.1. From 4.3 to 4.3.5, we then describe the components of the study: the participants, target forms, pre- and post-tests, treatment tasks, planning conditions and the procedure. From 4.4 to 4.4.5 we describe the analysis of the quantitative and qualitative data. 4.5 provides the results of the study.

In 4.6 we discuss the findings and finally, 4.7 reviews the pilot study and outlines the amendments required for the main study.

4.2 Quantitative, Qualitative and Mixed Research Methods

"Quantitative research involves data collection procedures that result primarily in numerical data which is then analysed primarily by statistical methods" (Dornyei, 2007, p. 24). This may involve the results of a language test or a survey questionnaire. These results are then usually analysed by computer software programs and converted into statistics upon which judgements can be made regarding the degree of significance. As research question one investigates learners' L2 speech which involves numerical data, as we saw in 2.4.2 that fluency measures can consist of 'number of syllables per minute' this investigation falls within quantitative research.

In contrast to quantitative research, "qualitative research involves data collection procedures that result primarily in open-ended, non-numerical data which is then analysed primarily by non-statistical methods" (Dornyei, 2007, p. 24). This may involve for example, recording speech from an interview which is then transcribed into written text for analysis. As research question two examines learners' planning strategies for oral tasks through post-task interviews, this investigation can be classified as qualitative research.

According to Dornyei (2007), qualitative research has recently had an important role to play in the field of applied linguistics because "almost every aspect of language acquisition and use is determined or significantly shaped by social, cultural and situational factors, and qualitative research is ideal for providing insights into such contextual conditions" (p. 36). For example, as we saw in Ortega (2005) in 3.3.7, L2 learners are influenced by their social and cultural backgrounds when planning for oral tasks. These insights into learners' planning strategies were made possible through qualitative data collection methods in the form of post-task interviews, and this research provides valuable information in helping to explain learners' linguistic performance in terms of fluency and accuracy. Richards (2003) points out that quantitative research, on the other hand, is "not designed to explore the complexities and conundrums of the immensely complicated social world that we inhabit" (p. 8). For example, tests and surveys can provide us with important information but if any unexpected results

occurred, quantitative research would be limited in explaining *how* or *why* the occurrence happened. The advantages of qualitative research enable us to probe unexpected outcomes by asking 'why'? and in doing so, "allows the researcher to conduct 'further research' straight away, thereby reaching a fuller understanding" (Dornyei, 2007, p. 40).

Although quantitative and qualitative research appears to be two separate paradigms, Dornyei (2007) is quick to point out that they are not mutually exclusive and that researchers can combine the best of both approaches to use, what is known as, "mixed methods research" (p. 24). This methodology "involves different combinations of qualitative and quantitative research either at the data collection or at the analysis levels". For example, it may involve studies that incorporate both questionnaires and interviews in order to improve the validity of the research as a whole by having one set of data that supports or justifies the results of the other. Creswell (1994) refers to the term "triangulation" (p. 174) when describing mixed methods research which involves using multiple forms of data collection to analyze specific outcomes, the advantage being that any bias shown in a particular source could be justified by another source. The findings of the study would then be based on the information as a whole thus offering credibility to the overall results.

Mixed methods research has developed relatively recently within the field of applied linguistics and there are calls for "more studies that combine qualitative and quantitative research methods, since each highlights "reality" in a different, yet complimentary way" (Lazaraton, 2005, p. 219) (cited in Dornyei, 2007, p. 44). As the present study is looking to analyse learners' L2 speech during task performance as well as investigating what learners did during planning, the use of quantitative and qualitative research respectively will serve to benefit the overall results of this study. The advantage being that the qualitative data i.e. learners' responses regarding their planning strategies would help shed light on the results of the quantitative data i.e. L2 performance in terms of fluency, accuracy and complexity. Creswell (1994) outlines three designs for mixed methods research. The first design, referred to as a "two-phase design" (p. 176), involves analysing data using one method, for example, qualitatively, then analysing data using another method, for example quantitatively. The second design, "dominant-less dominant" (p. 177), involves carrying out a study using a dominant method, for example, quantitative research whilst a certain aspect of the study that has less significance may involve qualitative research. Finally, a "mixed-

methodology design" involves using qualitative and quantitative research along many steps of the study. As the present study's main objective concerns research question one i.e. the effects of guided planning and task complexity on L2 oral development involving quantitative methods, whilst research question two serves as a supporting role involving learners' planning strategies using qualitative methods, we shall adopt a dominant-less dominant design.

4.2.1 Longitudinal research

Dornyei (2007) refers to longitudinal research as "the ongoing examination of people or phenomena over time" (p. 78). With regards to second language learning, this could involve tracking and analysing learners' progression of the L2 over a certain period in time. Menard (2002) defines longitudinal research according to the data and design of the study. For example, the data needs to be collected at two or more points in time and the comparison of the subjects in question are taken from the same sample group at the different points in time. This form of research differs from its counterpart 'cross-sectional research' which involves collecting data at just one point in time, as in the strategic planning studies we reviewed in 3.3.4. Menard (2002) notes the major advantages of longitudinal research is that it enables us "to describe patterns of change and to establish the direction (positive or negative and from Y to X or from X to Y)…..of casual relationships" (p. 3). This is of particular importance in terms of language learning, which in itself, is a process that occurs over time, thus Ortega & Iberri-Shea (2005) argue that longitudinal research is crucial in helping our understanding of *how* language learning works to a far greater degree than cross-sectional research which can only examine language performance at isolated points in time. For example, as research question two investigates learners' planning strategies over time, longitudinal research would enable data to be collected as learners plan for oral tasks at different time periods. Learners could be interviewed immediately after each performance which would enable them to easily recall their task performance. These results would therefore show us how language learning develops over time. Ortega & Iberri-Shea (2005) recommend incorporating qualitative and quantitative methods for longitudinal research because the strengths of each approach can compliment the findings of the overall study. As we saw in the last sub-section, this

appears to suit the purposes of the present study as we have already confirmed the use of a mixed methods approach therefore the most appropriate design would be to use a mixed-methods longitudinal design.

Given the apparent pedagogic value gained from longitudinal studies, why is there a lack of them within SLA research? For example, as we saw in 3.4.1, the only longitudinal planning study carried out to date was Bygate (2001b). As language learning is dynamic process, longitudinal studies would seem to be an ideal form of research, and Dornyei (2007) stresses the need for "many more longitudinal investigations in the field to explore the sequential patterns and the changes that occur" (p. 40). Yet the majority of language studies have instead performed cross-sectional research. Dornyei (2007) explains why this is the case by outlining the disadvantages of longitudinal research:

1 "Attrition" (p. 82). Due to the long-term nature of the research, participants may decide to withdraw during the process thus preventing the comparison of data at different points in time.
2 "Panel conditioning" (p. 83). Taking part in a longitudinal study may affect participants' behaviour and responses during the process to the point where it threatens the reliability of the data. For example, learners may "behave differently because they want to please the researcher whom they are getting to know better and better" (p. 83). Students may also "lose their inhibition about the data collection format" (p. 83). In relation to the present study, learners may become more relaxed each time they repeat a task and as a result, their performances may alter due to a reduction in their affective filters.
3 Samuda & Bygate (2008) also inform us of the extra cost and time that is required to carry out a long-term study as well as the extra data analysis involved which acts as a further deterrent.

In an attempt to address the issues involved with longitudinal studies, Dornyei (2007) recommends the following:

1 With regards to attrition, strategies can be used to prevent participants from pulling out of a study by providing small rewards. Attrition can also be minimised by reducing the length of a study. Dornyei (2007) argues that, within applied linguistics, a

longitudinal study may only need to last ten weeks, as in the case of Bygate's (2001b) study. To prevent attrition in this study, the researcher paid the students an hourly fee for their participation. Furthermore, the duration of the study was set at seven weeks, three weeks less than Bygate's (2001b) study, however it was decided that seven weeks would make it more convenient for teachers to replicate during a standard fifteen week university semester that may include examinations and mid-term breaks. The participants were initially informed about the length of the study, and were asked to consider their commitment before signing a contract which stipulated their willing participation (see appendix B for the participation consent forms).

2 Attempts were made to reduce the effects of panel conditioning by following Ortega (1999). For example, Ortega did not take part in task performances with the learners (the students performed them in dyads) and she only interacted with the participants during the interview sessions. Researcher biases were also avoided by not asking leading questions. For the purpose of this study, the researcher was required to participate with the students during task performance (to act as the listener) but no attempt was made to interact with the students on a personal level prior to task performance and during the interviews. In addition, as each student would wait in turn to see the researcher individually, there was no time available for personal discussion. Of course it is impossible to prevent participants from changing their behaviour over time. However, any negative affects that may have occurred from panel conditioning would have been recorded during the interviews and used as a factor to explain the quantitative results, thus showing the benefits of a mixed-methods approach as suggested by Ortega & Iberri-Shea (2005).

3 Although longitudinal studies require time and effort in order to collect data, the researcher was able to organise a schedule to allow the data collection to take place over a specified time period as well as finding time to analyse the data afterwards. In addition, as the data analysis involved the use of computer-aided qualitative and quantitative software, for example CLAN and SPSS (explained in more detail in the next chapter) this resulted "in considerable gains in efficiency. This frees up researcher time and helps to avoid data overload" (Dornyei, 2007, p. 265).

Having justified the mixed-methods longitudinal design for this study, we now turn our attention towards the design of the pilot study before examining each of its components in turn.

4.3 Experimental Design

In line with previous task-based developmental studies (Bygate, 2001b; Mackey, 1999), the pilot study incorporated a pre- post-test design that involved a pre-test, treatment, an immediate post-test followed by a two week delayed post-test. Schmitt (2010) explains that the benefits of such a design are that comparisons between the pre-test and the immediate post-test help "to determine the effect of the treatment" (p. 155) whilst comparisons between the pre-test and the delayed post-test "can demonstrate if long-term retention (i.e learning) has occurred" (p. 156). In addition, this study used a mixed factorial repeated-measures design. Dornyei (2007) notes that repeated-measures designs are commonly used within applied linguistics research and are relevant for quantitative longitudinal studies that involve analysing a sample of participants over certain points in time. The mixed factorial design consisted of within-subject and between-subject factors. The within-subject variable was testing with three levels: pre-test, immediate post-test, delayed post-test in which all the students performed the tests. Finally, the between-subject variable was planning time with three levels: guided planning, unguided planning, and no planning as the participants were split into three planning conditions.

4.3.1 The participants

Six intermediate and six upper-intermediate Japanese university learners of English participated in the study. The intermediate students were aged between 18 to 22 years old with a mean age of 19. Four students were male and two were female. They had been studying English for an average of 6.8 years and they were currently enrolled as English students at Ritsumeikan Asia Pacific University (APU) in Japan. Their English proficiency score averaged 470 TOEFL (test of English as a foreign language), varying from 462 to 483, and they were recruited from the researcher's intermediate level English language class at APU. In order to reduce the effects of panel conditioning

outlined in 4.2.1, the participants signed a consent form outlining an hourly payment for their services which stipulated that their involvement in the study would in no way influence their course grades.

The upper-intermediate learners were aged between 19 to 22 years old, with a mean age of 20. Four students were male and two were female. They also studied English at APU and were recruited from APU's low-advanced English language program. These participants had studied English for approximately 8.2 years, and they averaged a TOEFL score of 541, ranging from 453 to 835. For this advanced course, learners were expected to have a TOEFL score of 500, however, learners could graduate into the advanced class after completing the intermediate program thus classes could contain mixed ability which explains the varied range in this sample's proficiency. As a result of this sample's mean TOEFL score, it was decided they were to be classified as upper-intermediate learners, in line with a previous planning study (Kawauchi, 2005) that used upper-intermediate learners with a similar average TOEFL score of 545.

The purpose of recruiting learners of intermediate level proficiency were two fold, firstly it allows us to draw comparisons with the majority of previous planning studies which also used intermediate level learners (Ellis, 2009a). Secondly, the findings of this study enable us to see whether Japanese learners of intermediate proficiency could benefit from strategic planning to produce more positive results in terms of relativization compared to the beginner-level learners in Mochizuki & Ortega (2008) who mostly averaged a TOEFL score of "360-380" (p. 17) with a small number averaging "400-420" (p. 17).

A weakness of the TOEFL proficiency bands as indicators of proficiency is that they do not reflect a learner's speaking ability. Furthermore, in terms of previous studies, Ellis (2009a) points out that "many of the studies provide only very crude indicators of the proficiency level" (p. 491) and as result, he argues for future planning studies "to provide more explicit definitions of proficiency" (p. 493). As a result, the present study was interested in using an additional measure as an indicator of speaking performance. In order to achieve this, this study relied on The Common European Framework of References for Languages (CEFR) (2010) guidelines for proficiency levels. This framework classifies intermediate learners into groups: B1 and B2, the former is associated with intermediate level learners and the latter concerns upper-intermediate learners. In terms of speaking proficiency, the CEFR descriptions for B1 learners mention that they "can describe experiences and events, dreams, hopes and ambitions

and briefly give reasons and explanations for opinions and plans" (2010, p. 24). In terms of B2 oral proficiency, learners "can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party" (2010, p. 24). These descriptors appeared to be an appropriate match for the L2 oral ability of the respective intermediate and upper-intermediate level learners recruited for the study. Consequently, this study shall refer to the participants as B1 and B2 learners. The next sub-section describes the linguistic features chosen for the pilot study.

4.3.2 L2 target forms: OS and OO relative clause types and cognitive state verbs

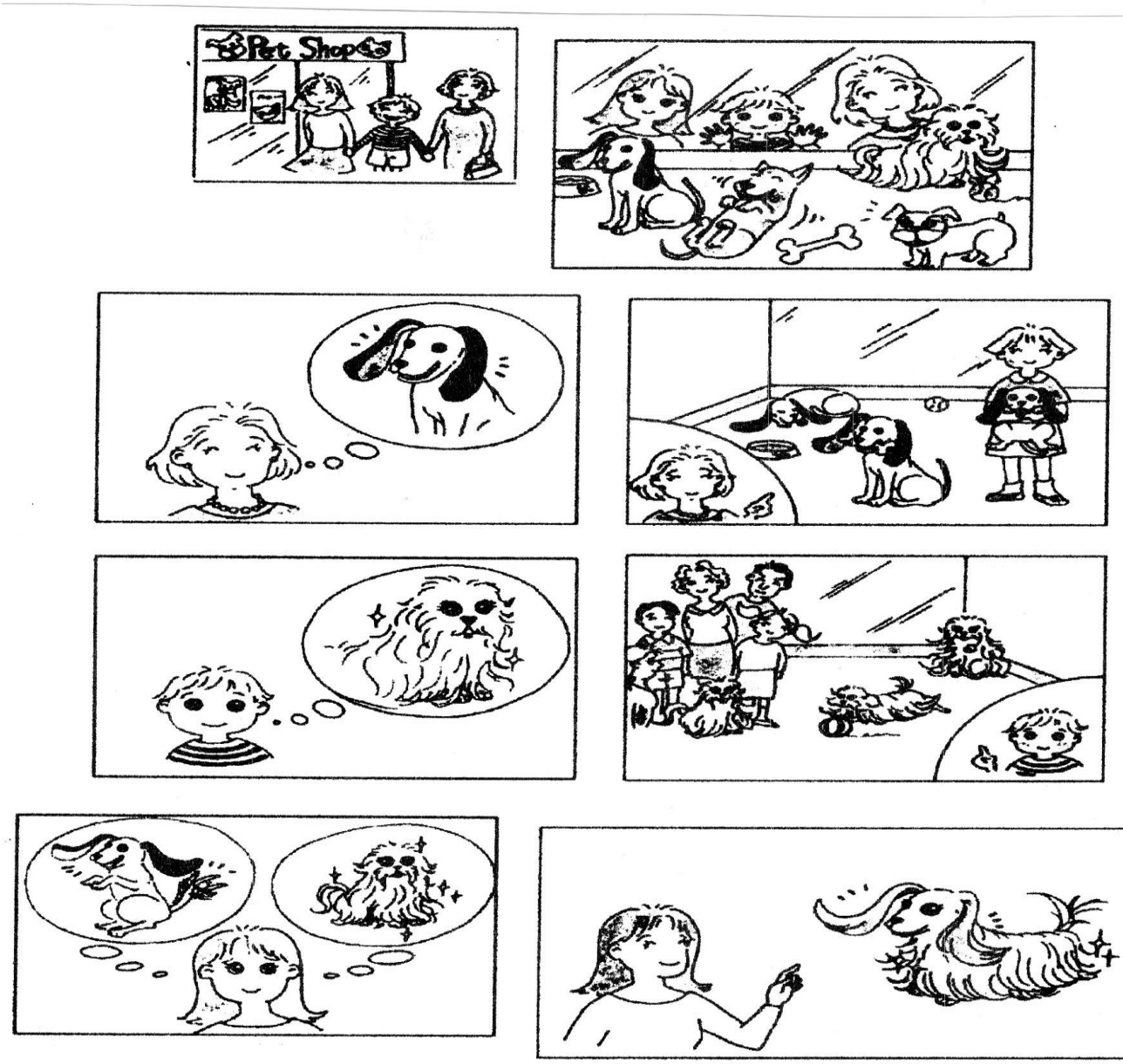
The linguistic forms targeted for the pilot study were English relative clauses (RCs) and the accompanying use of cognitive state verbs. In terms of RCs, two simple RC types used in Mochizuki & Ortega (2008) were targeted: object-subject (OS), *'she wants the doll which has black shoes'*, and object-direct object (OO) *'she wants the doll which the girl is watching'*. The results of Mochziuki & Ortega (2008) were disappointing in terms of the amount of RCs produced by the full sample which led Mochizuki & Ortega (2008) to conclude that the participants' beginner level proficiency was not high enough to benefit from the planning conditions. As a result, this study used the same RC types but with intermediate-level learners assuming that these learners would have sufficient explicit L2 knowledge to produce the RC types accurately. The results of the pilot study would then determine whether more complex RC types such as OPREP should be incorporated into the main study.

Cognitive state verbs were chosen in response to Robinson's (2007) study which reported their use when explaining the intentions of other people, for example, '*he thinks that…, he believes that…'*. These forms were considered compatible with RC production, for example, '*he thinks* that he likes the dog *which has long ears'*. Combining the use of cognitive state verbs with RCs is considered to be complex syntactic language for L2 learners to produce (Robinson, 2010) and as a result, this justified the use of the linguistic features as an appropriate linguistic challenge for intermediate level learners. We now turn to the assessments of the pilot study.

4.3.3 Pre- and post-tests

In order to measure L2 oral development, monologic narrative tasks were used. The pre- and post-test narratives were based on the task used in Mochizuki & Ortega (2008), which investigated RC production. For example, the pre-test was the task used in Mochizuki & Ortega (2008) and it consisted of eight pictures about a girl, her brother and mother who go to a pet-shop (see figure 7).

Figure 7. *Pre-test narrative (Mochizuki & Ortega, 2008, p, 36)*



Each pre- and post-test narrative contained the same number of obligatory cases of RCs as in Mochizuki & Ortega's (2008) task: seven. Mochizuki & Ortega (2008) provided examples of the seven RC cases used in their task (see appendix C). However, for the purpose of the present study, the RC examples were adapted to include the accompanying use of cognitive state verbs (see figure 8).

Figure 8: *Pre-test narrative seven obligatory RC contexts and accompanying use of cognitive state verbs (adapted from Mochizuki & Ortega, 2008, p. 35)*

Context 1 (OS): She thinks that she likes the dog which has long ears
Context 2 (OO): She wants the dog which the little girl has in her arms
Context 3 (OS): He believes that he likes the dog which has long hair
Context 4 (OO): He wants the dog which many people are watching
Context 5 (OS): She thinks that she likes the dog which has long ears
Context 6 (OS): She also thinks she likes the dog which has long hair
Context 7 (OS): She wants the dog which has long ears and long hair

For the purpose of this study, the narratives were expected to elicit cognitive state verbs because three pictures in each task contained thought bubbles from a character's head illustrating their feelings and this was intended to elicit language such as *'the girl <u>thinks</u> that she likes a dog…'* Each participant performed the task to the researcher who acted as the listener. The researcher asked each student to begin each narration in the present tense. For example, in pre-test, participants were asked to begin by saying, 'Today, kate and her family are at…'

Two different narratives were designed for the post-tests: one for the immediate and one for the delayed post-test. However, the pre- post-test narratives were designed to be similar in terms of cognitive complexity. As mentioned in 3.4.3, cognitive complexity depends of resource-directing factors *and* resource dispersing factors. The former relates to the linguistic demands of a task. In terms of the oral narratives, each version was the same along resource-directing dimensions because each task contained seven obligatory contexts of the same RC types and accompanying use of cognitive state verbs. To avoid using the same tasks however, the narratives were different in terms of storyline. For example, the intermediate narrative test involved a boy and his

parents going to a shoe shop, whilst the delayed narrative test involved three sisters going to a garden centre to buy plants (see appendix D).

Finally, the pre- and post-test narratives did not involve planning time in order to reflect natural conditions for language use (Robinson, 2005). Thus, it was not expected that the participants would produce many instances of the targeted forms during the pre-test but the planning groups were expected to produce the targeted forms during the post-tests after receiving their respective task sequencing treatment. In addition, the prevention of planning time would create conditions to test learners' implicit knowledge, which as we discussed in 3.5.2, involves unconscious, fluent processing of language which can be evaluated using unplanned testing conditions. As a result, the pre- and post-test narratives were of similar complexity along 'resource-dispersing' dimensions which relate to the availability of planning time. As the tests involved no planning time, they were each considered complex because learners had no time prepare prior to performance.

Finally, the advantage of using narratives based on Mochizuki & Ortega (2008)'s task was that it allowed us to draw comparisons to see whether learners of intermediate proficiency could have more success in producing the same RC types compared to the beginner learners in Mochizuki & Ortega's (2008) study. The next sub-section describes the treatment tasks used to develop learners' L2 speech.

4.3.4 The treatment tasks

Six narratives were designed for the treatment sessions, again based on the narrative used in Mochizuki & Ortega (2008) (see appendix E). However, these tasks differed in terms of content *and* cognitive complexity. In terms of content, they contained different storylines and characters. For example, narrative three involved a boy and three brothers going to a toy shop. In terms of cognitive complexity, the treatment narratives were designed and sequenced to increase in complexity according to the claims of Robinson's (2003) Cognition Hypothesis which was described in 3.4.3. This involved sequencing the tasks so they increased in complexity along resource-directing dimensions by increasing the linguistic demands of the tasks. This was achieved by adding obligatory contexts of the targeted RC types within each narrative. For example, narratives one and two contained seven RC contexts, in line with the pre-test, however

narratives three and four contained nine instances whilst narratives five and six contained ten RC contexts (see table 9).

Table 9. *Obligatory cases of relative clauses*

Target Forms	Pre-test Narrative	Task treatment (increasing intentional reasoning demands)			Immediate Post-test Narrative	Delayed post-test narrative
		Narrative 1 & 2	Narrative 3 & 4	Narrative 5 & 6		
Relative Clauses	7	7	9	10	7	7

The treatment narratives therefore increased along resource-directing dimensions through increased intentional reasoning demands which, as we saw in 3.4.3 requires the learner to explain the actions of other people. For example, the narratives contained additional instances of characters thinking about something, in the form of thought bubbles, and this was intended to elicit further production of RCs and accompanying use of cognitive state verbs. We now outline the procedure of the study.

4.3.5 The procedure

Data collection for the study took place in a recording studio at APU during June-July 2011. The six intermediate and six upper-intermediate learners were randomly split into three pairs per proficiency level: guided planning (GP), unguided planning (UP) control group/no planning (NP) (see table 10).

Table 10. *Participant pairs*

Guided Planners (GP)	2 intermediate learners	2 upper-intermediate learners
Unguided Planners (UP)	2 intermediate learners	2 upper-intermediate learners
Control Group (NP)	2 intermediate learners	2 upper-intermediate learners

In week one, the pre-test was carried out. In week two the treatment sessions began for the GP and UG planners. Each pair was initially provided with a ten minute instruction workshop where the researcher explained the planning conditions. Each pair was then allocated ten minutes planning time prior to each task in which to make notes. The length of planning time was based on the majority of previous planning studies that used ten minutes planning time (Ellis, 2009). The GP pairs received guidance during their planning time in the form of note-sheets that contained written examples of the targeted RC types and cognitive state verbs. The note-sheets were adapted from the guided planning note-sheets used in Mochizuki & Ortega (2008) and Sangarun (2005) (see appendix F). The GP pairs were instructed to read the examples provided and try to use the grammar in their story. The UP pairs received no guidance during planning, however, following Yuan & Ellis (2003) the learners were briefly informed to think about organization, language and content. Finally, the GP and UP planners were told they could not use their notes during the task. After completing a task, each participant was interviewed one-on-one with the researcher in a separate room regarding their planning strategies. Weeks three and four of the treatment sessions were repeated in the same format as week two. As all the treatment sessions involved ten minutes planning time, the tasks were considered simple in terms of resource-dispersing dimensions which relates to the availability of planning time. The control groups did not perform any tasks during the treatment sessions. At week five, all the participants performed the immediate post-test narrative. Finally, at week seven, all the participants performed the delayed post-test (see table 11).

Table 11. *Pilot Schedule*

Pairs	Pre-test	Treatment (increasing intentional reasoning demands)			Immediate Post-test	Delayed Post-test
	Week1	Week 2	Week 3	Week 4	Week 5	Week 7
Guided Planners	Narrative	Narrative 1 & 2 & post-task Interview	Narrative 3 & 4 & post-task Interview	Narrative 5 & 6 & post-task Interview	Narrative	Narrative
Unguided Planners	Narrative	Narrative 1 & 2 & post-task Interview	Narrative 3 & 4 & post-task Interview	Narrative 5 & 6 & post-task Interview	Narrative	Narrative
Control Group	Narrative				Narrative	Narrative

Having discussed the materials and the procedure of the pilot study, the next sub-section describes the analysis measures.

4.4 Data Analysis

As discussed in 4.2, research question one concerned the analysis of L2 oral development using quantitative methods. The target variables used for comparison, otherwise known as dependent variables (Dornyei, 2007) were the measures used for complexity, accuracy and fluency. CAF was analysed by comparing the planning pairs and the control group's performance at the pre-tests in week one against the immediate post-test at week five and the delayed post-test at week seven. L2 development was then determined by the extent of the pre- post-test gains. We now turn to examine each of the measures used to analyse CAF then we provide a description of the qualitative measures used to investigate research two which concerns the learners' planning strategies.

4.4.1 Accuracy Measures

As mentioned in 2.4.3, our definition of accuracy relates to "the extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm (i.e. usually the native speaker). Thus, deviations from targetlike performance

would be considered errors" (Housen et al. 2012, p. 4). In addition, accuracy also relates to "*appropriateness* and *acceptability*" (p. 4). For the purpose of this study, deviations from targetlike performance relate to grammatical errors *and* communicatively inadequate use of the targeted forms: OS and OO RC types as well as the accompanying use of cognitive state verbs. In other words, if a learner produced an OS RC type that was grammatically correct but it did not reflect the context of the storyline it would be considered inaccurate. It was therefore necessary to use a measure that could gauge grammatical errors relating to OS and OO RC types and the accompanying use of cognitive state verbs. As the participants in the present study were of similar intermediate proficiency to previous studies (for example, Yuan & Ellis, 2003, Foster & Skehan, 2005), it was decided that similar accuracy measures should be used to enable comparisons with other studies. As we saw in 2.4.4, studies which have investigated syntactic accuracy such as Yuan & Ellis (2003) used 'error-free clauses', specifically, the percentage of clauses produced that do not contain any lexical, syntax or morphology error. A typical lexical error could include "I was waiting you" (p. 14). As a result, the present study measured grammatical accuracy according to the percentage of '*error-free relative clauses per relative clause'*. This involved counting all the relative clauses that contained no grammatical errors and dividing them against the total number of relative clauses produced per narrative as follows:

$$\frac{\text{Number of error-free relative clauses}}{\text{Number of relative clauses}} *100$$

An example of an error-free RC could be 'the girl likes the dog *which has long ears'*, in this case, only the RC would be measured. A typical grammatical error could therefore be *'*the dog *which <u>have</u> long ears'.*

In terms of the remaining part of our accuracy definition: communicatively adequate use of the targeted forms, learners' accuracy was based on the context in which they used the forms. In other words, in line with Mochizuki & Ortega (2008), the learners were required to produce seven obligatory contexts of RCs that reflected the storyline, thus if an RC was produced that was grammatically correct but communicatively inadequate i.e. it did not relate the storyline it would be excluded from the analysis. For example, describing a picture which contains a dog with long hair but commenting *'the cat which has short hair'*. In terms of over-use of the forms, repeated

RCs were excluded from the analysis. For example, as in the underlined structures 'he likes the dog *which has black hair, which has black hair*' therefore if a learner produced twelve RCs during a narration, they would only be graded on the seven obligatory contexts as shown in figure 8.

As we saw in 2.4.4 previous studies have also used multiple measures for CAF, therefore an additional measure for accuracy was used 'percentage *error-free relative clauses per AS-unit'*. An AS-unit is defined as "an independent clause or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster, Tonkyn & Wigglesworth, 2000, p. 365). Although the definition of an AS-unit is very similar to the definition of a t-unit or a c-unit which was described in 2.4.6, as they all consist of an independent clause with any subordinate clauses attached, an AS-unit was chosen because it "allows for the inclusion of sub-clausal units which are common in speech" (p. 366). Sub-clausal units comprise of elliptical utterances which are shortened phrases that can be inferred into full clauses based on the pragmatic meaning of the situation. Foster, Tonkyn & Wigglesworth (2000, p. 366) provide an example:

A: How long you stay here
B: Three months

Although these phrases contain grammatical errors, they have pragmatic meaning within the context and therefore each utterance would be considered an AS-unit. However, repeated clauses and incomplete clauses were excluded from the analysis. For example, as in the underlined structures 'he likes the dog *which has black hairs, which has black hair*.' Thus the following formula was used:

Number of error-free relative clauses
Number of AS-units *100

Accurate use of cognitive state terms was measured in a similar fashion to relative clauses: *'percentage of error-free cognitive state verbs per AS-unit'* and '*percentage of error-free cognitive verbs per cognitive state verb'*. For the purpose of the pilot study, cognitive state verbs concerned the underlined verbs on the guided planning note-sheets, for example: '*She thinks that she likes the dog which looks friendly*, '*She wants the doll which has black shoes*'. Typical errors involving the use of

cognitive state verbs included incorrect subject-verb agreement, for example, '*He think that he…'*, '*He thinks like dog…'* The analysis involved counting all the cognitive state verbs that were used accurately and dividing them against the total number of AS-units for the former measure, and total number of cognitive state verbs produced for the latter measure as follows:

$$\frac{\text{Number of error-free cognitive state verbs}}{\text{Number of AS units}} *100$$

$$\frac{\text{Number of error-free cognitive state verbs}}{\text{Number of cognitive state verbs}} *100$$

Further examples and formulas for measuring accuracy will be explained in more detail in the next chapter (5.7.2).

4.4.2 Syntactic complexity measures

In 2.4.5 we defined syntactic complexity as:

> a language feature or system of features is seen as complex if it is somehow costly or taxing for language users and learners, particularly in terms of the mental effort or resources that they have to invest in processing or internalizing the feature(s). (Bulte & Housen, 2010, p. 23)

For the purpose of this study, syntactic complexity refers to English RCs which are a grammatical feature known for its cognitive difficulty in L2 oral production and development for Japanese learners. It was therefore important to use a measure that reflected our definition. Following Mochizuki & Ortega (2008), the 'amount of relative clauses per AS-unit' was used, as we know from 3.5.4 relative clauses are known for their difficulty with Japanese learners due to L1 and L2 differences in the form. Following Mochizuki & Ortega (2008), non-target-like use of RCs was accepted provided the participant used a relative pronoun. However, in line with our accuracy measure, repeated RCs were excluded from the analysis in order to prevent over-use of the form, for example, as in the underlined structures 'he likes the dog *which has black hair*, *which has black hair*'. In addition, following Mochizuki & Ortega (2008), one

more syntactic complexity measure was used: 'dependent clauses per AS-unit'. Although this measure did not relate as accurately to our definition of syntactic complexity, it was used in order to draw comparisons with the 'relative clauses per AS-unit' measure. However, unlike Mochizuki & Ortega (2008), this study used an AS-unit instead of a t-unit to keep consistency with our accuracy measures. Dependent clauses, otherwise known as subordinate clauses consist of a verb and one other clause feature, for example, a subject or object. Foster, Tonkyn & Wigglesworth (2000) provide an example, "it is my hope :: to study crop protection (2 clauses, 1 AS-unit)" (p. 366). The underlined clause refers to the dependent clause. Conventional phrases such as '*hello', 'that's all'* were excluded from the analysis. Further examples relating to the coding of syntactic complexity will be explained in the next chapter 5.7.3.

4.4.3 Fluency measures

In 2.4.1, we defined fluency as "the rapid, smooth, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (adapted from Lennon, 2000, p. 26). We then identified a measure that was a suitable match for our definition known as 'pruned speech rate' which relates to "the average number of syllables produced per minute of pruned speech, i.e. speech from which repetitions, false starts and other performance features have been excluded" (Levkina & Gilabert, 2010, p. 182). In 2.4.2 we briefly justified the choice of this measure, for example, syllables per minute relates to rapid speech. In 5.7.4 this measure is justified in detail however for the purpose of this pilot study 'other performance features' related to self-corrections, L1 use and incomprehensible language. Thus examples of language that was omitted included the underlined structures for repetitions: *'the the man is sad'*, false starts: *'the ca the dog'*, self-corrections: *'The woman err the man went to the shop'*, whilst L1 use included Japanese language.

As noted in 2.4.2, in order to measure fluency in terms of the final part of our definition 'under the temporal constraints of on-line processing' required a test that did not involve planning time. This was achieved by preventing strategic planning time during the pre- and post-test narrations which forced the learners to engage in 'on-line processing.'

For the purpose of the pilot study, an additional similar measure was chosen from Gilabert (2007a) 'unpruned speech rate' which involves the total number of syllables produced per minute. The use of multiple measures would allow us to draw comparisons with learners' performance in previous studies such as Yuan & Ellis (2003) and would enable us to choose a suitable measure for the main study. Pruned speech was calculated as:

Total number of syllables (excluding repetitions, self-corrections, false-starts, L1 use and incomprehensible language)
Total number of seconds x 60

Unpruned speech was calculated as follows:

Total number of syllables
Total number of seconds x 60

In terms of total speaking time, syllables were recorded from the participant's use of '*Today…'* and finished at the end of the narration. Phrases such as '*that's all', 'finished'* were excluded. Further examples relating to the coding of fluency will be explained in the next chapter 5.7.4.

4.4.4 Post-task interview questions

In order to answer research question two, namely how learners plan for oral tasks that increase in complexity over time, a qualitative analysis was carried out in the form of post-task interviews. The GP and UP pairs were interviewed about their planning strategies after they had performed a narrative in weeks two, three and four of the treatment sessions. Each interview lasted between six and ten minutes.

The interview questions were taken from Ortega (1999) who also investigated learners' planning strategies for oral tasks (appendix G contains all the questions used). Further justification for the choice of questions will be explained in the following chapter (5.6.1). The order of the main questions asked in the present study was as follows:

Q1 How did you plan?

Q2 What was your focus when you prepared the story?

Q3 Did you think about vocabulary, grammar, how to organize your story, or something else?

Q4 Did you plan differently compared to last time? In what way?

Questions three and four were designed for the purpose of this study. Question three was adapted from a questionnaire in Yuan & Ellis (2003) and was used in an attempt to prompt learners to think about form and meaning if they did not provide responses about their planning strategies from questions one and two. Finally, as research question two investigates learners planning strategies over time, question four was used during weeks three and four of the treatment in order to compare planning strategies from previous weeks. Finally, the researcher was required to conduct the interviews in English as it was not possible to conduct the interviews in the learners' L1. It was therefore important that the questions were comprehensible for intermediate level learners. The questions were therefore pre-piloted on two intermediate learners at APU who did not participate in the study but were considered to have B2 level speaking ability. Both students were interviewed after performing the pre-test narrative with ten minutes guided planning time. The students were able to comprehend the questions and were also able to describe how they planned for the tasks in English. As a result, no changes were made to the interview questions.

4.4.5 Instrument analysis, transcriptions and coding

The small sample size of each pair (n = 2) prevented any inferential statistical analysis from being carried out as small sample sizes violate the assumptions of most statistical analyses such as ANCOVA (Field, 2009). In addition, analyses involving effect sizes using for example, Cohen's *d*-value were also not possible as a minimum sample (n = 8) is required per group (Cohen, 1988). As a result, descriptive statistics were calculated to

provide mean values of each pair's pre-test and post-test scores. The analysis then involved comparing the pre- and posts-test mean scores for developmental gains in accuracy, syntactic complexity and fluency for each pair of learners. The pre- and post-test narrative performances were recorded by the researcher then transcribed onto the software program CLAN for the coding of fluency, accuracy and complexity (see appendix H for the narrative transcriptions). The use of CLAN will be justified in the next chapter. In terms of research question two, all interviews during weeks two, three and four of the treatment were recorded by the researcher, transcribed onto MS word and then analysed for specific patterns or differences regarding each pairs' planning strategies (see appendix I for the interview transcriptions). The following section outlines the results of the study.

4.5 Pilot study results

The analysis begins by comparing the differences between the pre-tests and posts-tests for developmental gains in accuracy. This was done by analysing the descriptive statistics of the B1 and B2 guided planners' production of RCs against their respective unguided planners and control groups. Finally, the results of the B1 and B2 learners were compared to see which proficiency level improved the most in RC development as a result of the treatment. This procedure was then repeated for the additional measure of accuracy: cognitive state verbs, as well as syntactic complexity and fluency.

4.5.1 Results of hypothesis one: effects on accuracy of OS and OO RC types

According to hypothesis one, guided planning and task complexity will facilitate L2 oral development to a greater extent than unguided planning and task complexity, and the control group in terms of accurate use of OO and OS RC types. Hypothesis one was confirmed.

The percentage results of the two accuracy measures used 'error-free relative clauses per AS-unit' and 'error-free relative clauses per relative clause' for the B1 and B2 learners are displayed in table 12. The B1 participants consisted of two B1 guided planners (B1GP), two B1 unguided planners (B1UP), and two B1 control learners

(B1CP). The B2 participants consisted of two B2 guided planners (B2GP), two B2 unguided planners (B2UP) and two B2 control learners (B2CP). Each pair contains their mean score. ‘Pre–test immediate post-test difference’ represents the difference between the pre-test and the immediate post-test scores, whilst the ‘pre–test delayed post-test difference’ shows the difference between the pre-test and the delayed post-test scores.

Table 12. *Descriptive results of the two accuracy RC measures*

	Percentage of error-free relative clauses per AS-unit					Percentage of error-free relative clauses per Relative clause				
Pairs (n = 2)	Pre-test	Immediate Post-test	Delayed Post-test	**Pre-test Immediate Post-test Difference**	**Pre-test delayed post-test Difference**	Pre-test	Immediate Post-test	Delayed Post-test	**Pre-test immediate post-test difference**	**Pre-test Delayed post-test Difference**
B1 GP	4.55	6.25	42.78	**1.71**	**38.24**	25.00	8.34	87.50	**-16.67**	**62.50**
B1 UP	6.25	0.00	19.88	**-6.25**	**13.63**	25.00	0.00	50.00	**-25.00**	**25.00**
B1 CP	0.00	0.00	5.56	**0.00**	**5.56**	0.00	0.00	25.00	**0.00**	**25.00**
B2 GP	4.17	22.22	57.50	**18.06**	**53.34**	10.00	28.57	70.84	**18.57**	**60.84**
B2 UP	18.75	0.00	5.00	**-18.75**	**-13.75**	50.00	0.00	0.00	**-50.00**	**-50.00**
B2 CP	8.34	0.00	0.00	**-8.34**	**-8.34**	50.00	0.00	0.00	**-50.00**	**-50.00**

Comparing the mean scores showed a clear effect for the B1 and B2 guided planners compared to their respective unguided planners and the control learners using both measures. Apart from the B1GP pre-test immediate post-test difference using the 'error-free relative clauses per relative clause' measure (M = -16.67), there were clear improvements in the accuracy of the guided planners compared to the unguided planners and the control groups. However, the B2GP learners showed the greatest gains in error-free relative clause production across both measures. We can therefore conclude that guided planning and task complexity produces greater gains in RC accuracy with the B1 and B2 learners compared to unguided planning and task complexity and the control groups which remain over time.

4.5.2 Results of hypothesis two: effects on accuracy of cognitive state verbs

According to hypothesis two, guided planning and task complexity will facilitate L2 oral development to a greater extent than unguided planning and task complexity, and the control groups in terms of accurate use of cognitive state verbs. Hypothesis two was partly confirmed. The percentage results of the two measures used for cognitive state verbs 'error-free cognitive state verbs per AS-unit' and 'error-free cognitive state verbs per cognitive state verb' for the B1 and B2 learners are displayed in table 13.

Table 13. *Descriptive results of the two accuracy cognitive state verb measures*

	Percentage of error-free cognitive state verbs per AS-unit					Percentage of error-free cognitive state verbs per Cognitive state verb				
Pairs (n = 2)	Pre-test	Immediate Post-test	Delayed Post-test	**Pre-test Immediate post-test difference**	**Pre-test Delayed post-test difference**	Pre-test	Immediate Post-test	Delayed Post-test	**Pre-test immediate post-test difference**	**Pre-test delayed post-test difference**
B1 GP	38.18	67.62	67.22	**29.44**	**29.04**	83.36	91.67	100.00	**8.33**	**16.67**
B1 UP	8.825	8.83	4.17	**0.00**	**-4.66**	30.00	30.00	16.67	**0.00**	**-13.34**
B1 CP	5	10.10	0.00	**5.10**	**-5.00**	12.50	33.36	0.00	**20.84**	**-12.50**
B2 GP	29.77	79.17	100.00	**49.40**	**70.23**	64.29	88.89	100.00	**24.61**	**35.72**
B2 UP	0.00	15.00	20.56	**15.00**	**20.56**	0.00	50.00	62.50	**50.00**	**62.50**
B2 CP	16.67	0.00	6.25	**-16.67**	**-10.42**	25.00	0.00	8.36	**-25.00**	**-16.67**

These results showed a clear effect for the B1 and B2 guided learners compared to the unguided planners and the control groups using ‘error-free cognitive state verbs per AS-unit’. However, in terms of the B1 learners, under the ‘error-free cognitive state verbs per cognitive state verb’ measure, the B1 control group showed the largest pre-test immediate post-test gain (*M* = 20.84). Although the B1 guided planners showed the largest pre-test delayed post-test gain (*M* = 16.67). In terms of the B2 learners, the unguided planners produced the greatest gains from the pre-test to the immediate post-test (*M* = 50) as well as from the pre-test to the delayed post-test (*M* = 62.5). We can therefore conclude that guided planning and task complexity produces greater accuracy gains in the use of cognitive state verbs with the B1 and B2 learners compared to unguided planning and task complexity and the control group using ‘error-free cognitive state verbs per AS-unit’. However, this is not confirmed when using ‘error-free cognitive state verbs per cognitive state verb’ due to the larger gains from the B1 control group, and the B2 unguided planners.

4.5.3 Results of hypothesis three: effects on syntactic complexity

According to hypothesis three, guided planning and task complexity will facilitate L2 oral development to a greater extent than unguided planning and task complexity, and the control group in terms of syntactic complexity. Hypothesis three was confirmed. Table 14 displays the two complexity measures used; ‘relative clauses per AS-unit’ and ‘dependent clauses per AS-unit’ for the B1 and B2 learners.

Table 14. *Descriptive results of the two syntactic complexity measures*

	Number of relative clauses per AS-unit					Number of dependent clauses per AS-unit				
Pairs (n=2)	Pre-Test	Immediate Post-test	Delayed Post-Test	**Pre-test Immediate Test Difference**	**Pre-test Delayed Test Difference**	Pre-Test	Immediate Post-test	Delayed Post-test	**Pre-test Immediate Test Difference**	**Pre-test Delayed Test Difference**
B1 GP	0.39	0.64	0.47	**0.25**	**0.08**	0.53	1.12	0.86	**0.59**	**0.33**
B1 UP	0.33	0.14	0.48	**-0.18**	**0.15**	0.46	0.3	0.76	**-0.16**	**0.31**
B1 CP	0.23	0.37	0.44	**0.14**	**0.21**	0.72	0.64	0.84	**-0.08**	**0.12**
B2 GP	0.27	0.68	0.80	**0.40**	**0.52**	0.7	0.97	1.19	**0.27**	**0.49**
B2 UP	0.37	0.00	0.22	**-0.37**	**-0.15**	0.57	0.15	0.54	**-0.42**	**-0.03**
B2 CP	0.16	0.25	0.23	**0.08**	**0.06**	0.92	0.8	1.15	**-0.12**	**0.23**

These results showed a clear effect for the B1 and B2 guided planners. Gains are reported for the B1 guided planners across both measures and they are greater than the B1 unguided learners and the control group with the exception of the pre-test delayed post-test gain using 'relative clauses per AS-unit'. With regards to the B2 learners, the B2 guided planners produced greater gains in complexity compared to the unguided planners and the control group using both measures. Thus, the B2GP learners appear to show the greater gains in complexity using both measures. We can therefore conclude that guided planning and task complexity produces greater gains in syntactic complexity with the B1 and B2 learners compared to unguided planning and task complexity and the control groups which remain stable over time.

4.5.4 Results of hypothesis four: effects on fluency

According to hypothesis four guided planning and task complexity will facilitate L2 oral development to a lesser extent than unguided planning and task complexity, and a control group in terms of fluency. Hypothesis four was partly confirmed. The results of the two fluency measures used: unpruned speech rate A (total number of syllables divided by total number of seconds multiplied by 60) and pruned speech rate B (total number of syllables excluding repetitions, false starts, self-corrections, L1 use and incomprehensible language) for the B1 and B2 learners are shown in table 15.

Table 15. *Descriptive results of unpruned and pruned fluency speech rate measures*

	Fluency rate a (unpruned speech): number of syllables per minute					Fluency rate b (pruned speech): number of syllables per minute excluding repetitions, false-starts, self-corrections, L1 use and incomprehensible language				
Pairs (n=2)	Pre-test	Immediate Post-test	Delayed Post-test	**Pre-test Immediate Post-test Difference**	**Pre-test delayed post-test difference**	Pre-test	Immediate Post-test	Delayed Post-test	**Pre-test immediate post-test difference**	**Pre-test delayed post-test difference**
B1 GP	83.13	93.77	88.33	**10.64**	**5.2**	71.16	80.72	82.26	**9.56**	**11.1**
B1 UP	77.66	76.56	87.07	**-1.1**	**9.41**	52.8	57.31	67.53	**4.51**	**14.73**
B1 CP	94.5	95.63	91	**1.13**	**-3.5**	70.79	72.07	69.08	**1.28**	**-1.71**
B2 GP	122.81	104.46	122.14	**-18.35**	**-0.67**	98.27	91.86	109.4	**-6.41**	**11.13**
B2 UP	89.85	103.33	110.1	**13.48**	**20.25**	59.37	93.68	91.61	**34.31**	**32.24**
B2 CP	90.56	154.21	112.99	**63.65**	**22.43**	76.58	121.84	92.8	**45.26**	**16.22**

Guided planning and task complexity produced varied results for fluency depending on the measures used. For the B1 learners, the guided planners showed the greatest gains from the pre-test to the immediate post-test using both measures. However, the unguided planners produced the largest gains from the pre-test to the delayed post-test using both measures. In terms of the B2 learners, there was a negative effect for guided planning using both measures, with one positive effect shown from the pre-test to the delayed post-test using the 'rate b' measure (M = 11.13). More substantial gains were shown with the unguided planners and the control group using both measures with the control group showing the largest gains. As a result, we can conclude that guided planning and task complexity produces greater gains in fluency from the pre-test to the immediate post-test for the B1 learners compared to unguided planning and task complexity and the control group. However, this is not confirmed at the delayed post-test. In terms of the B2 learners, guided planning and task complexity does not produce gains in fluency compared to unguided planning and task complexity and the control group using both measures.

Having discussed the quantitative results for research question one, the next sub-section describes the qualitative results for research question two.

4.5.5 Results for hypothesis five: guided planners' strategies

According to hypothesis five, the intermediate and upper-intermediate guided planners would initially focus on form then as tasks increase in complexity they would gradually attend more towards meaning over time. Hypothesis five was partly confirmed.

The analysis begins by looking at the post-task interview responses of the two B1 guided planners (B1GP 1) and (B1GP 2) as they progressed from simple to complex tasks during weeks two, three and four of the treatment. The above steps were then repeated for the B2 guided planners (B2GP 1) and (B2GP 2). The responses from all the participants were not corrected.

The planning strategies of B1 guided planners appeared to fluctuate through time (see table 16). Initially, they appeared to focus on organization so they could communicate the main idea of the story. However, as the tasks increased in complexity the learners seemed to focus more on grammar due to the extra details in the task and

attempts were made to use the target grammar. At the end of the treatment, they maintained their attention on grammar whilst also attempting to speak more fluently.

Table 16. *B1GP strategies*

Week 2	B1GP 1	"My focus was I just tried to describe each pictures well I just tried to write a lot without thinking grammar"
	B1GP 2	"Er I was thinking about how to organize the story"
Week 3	B1GP 1	"I was trying to use these sentences like she thinks that She likes the doll which…"
	B1GP 2	"I was very conscious about grammar because if I don't Use the correct grammar then my speaking will be chaos"
Week 4	B1GP 1	"I tried to use this grammar and try to describe each picture separately"
	B1GP 2	"I think I describe a lot than before tried to describe more details, I tried not to say er er er"

The strategies of the B2 guided planners appeared to be more consistent over time. They began planning by focusing on grammar then as the weeks progressed, they appeared to maintain their attention on grammar in an attempt to the use the forms accurately. They also seemed to focus their attention on vocabulary as the treatment progressed (see table 17).

Table 17. *B2GP strategies*

Week 2	B2GP 1	"I was trying to think about vocabulary ah no no sorry grammar is most important"
	B2GP 2	"I think about the grammar which I learned today"
Week 3	B2GP 1	"I was trying to make this type of grammar make sentences with this type of grammar"
	B2GP 2	"I think that's erm like grammar and how to organize story"
Week 4	B2GP 1	"Definitely vocabulary because I didn't know the names."
	B2GP 2	"Especially on grammar structure and as I said before tried to use the transition words"

After reviewing the planning strategies of the B1 and B2 guided learners, it appears that hypothesis five has been partly confirmed. Initially, the B1GP learners did not focus on form, however the B2GP learners did. As the sessions progressed, both the B1GP and B2GP learners attended to grammar whilst there is evidence towards the end of planning towards fluency from the B1GP learners.

4.5.7 Results for hypothesis six: unguided planners' strategies

According to hypothesis six, the unguided planners would initially focus on meaning, then as tasks increased in complexity these learners would gradually attend more towards form over time. Hypothesis six was not confirmed. The planning strategies of the two B1 unguided learners (B1UP 1) and (B1UP 2) appeared to remain fairly constant through time. They began by focusing on grammar to describe all the features of the characters in the pictures and organising their stories. They then continued this planning strategy throughout the treatment (see table 18).

Table 18. *B1UP strategies*

Week 2	B1UP 1	"I was careful for the grammar and er all the time but I don't think it was right"
	B1UP 2	"Er organise the story. I tried to use same words and same grammar"
Week 3	B1UP 1	"This time I I the most important thing is organise the story"
	B1UP 2	"Focus was er sisters wanted to buy similar dolls so I I focus the word"
Week 4	B1UP 1	"Er I think it is important for me to to speak rightly or to pay attention to grammar."
	B1UP 2	"Er I yeah I thought about vocabulary not vocabulary, grammar"

In terms of the B2 unguided planners (B2UP) they also appeared to use similar planning strategies over time. They began by planning to explain the details of the pictures then as they attempted more complex tasks they appeared to focus their attention on explaining the story in more detail which involved attention to grammar (see table 19).

Table 19. *B2UP strategies*

Week 2	B2UP 1	"Er I was thinking about story ofcourse I have to focus on the story and then I firstly grammar"
	B2UP 2	"My focus is the different erm different about monkey"
Week 3	B2UP 1	"I wanted to explain clearly so firstly I focus on the grammar"
	B2UP 2	"I understand the picture and try to explain more detail"
Week 4	B2UP 1	"This time I focused on not vocabulary but grammar for example er yeah I was used other grammar"
	B2UP 2	"Ah erm I I focus the cars characteristics"

After reviewing the planning strategies of the B1 and B2 unguided planners it appears that hypothesis six was not confirmed. The B1 and B2 unguided learners tended to plan by explaining there stories in detail, focusing on form not meaning and these strategies remained largely unchanged throughout the treatment.

4.6 Discussion

The main purpose of this pilot study was to examine the extent to which guided planning and task complexity facilitates L2 oral development in terms of accurate use of OS and OO RC types, cognitive state verbs, fluency and syntactic complexity of B1 and B2 Japanese learners of English. The results of the study showed that guided planning and task complexity facilitated L2 oral development with regards to the accuracy of OS and OO RCs and cognitive state verbs, syntactic complexity and fluency for this sample of B1 learners and B2 learners. This treatment also appeared more powerful than unguided planning and task complexity, and the control groups, except in the case of fluency and cognitive state verb development for the B2 guided planners.

In terms of accuracy, although guided planning and task complexity facilitated gains in the targeted forms, the all-or-nothing binary feature of the measures used, for example, 'percentage of *error-free* relative clauses per relative clause' meant that not many cases were produced by the B1 and B2 learners, for example, the pre- immediate post-test mean gain of the B1 guided learners was (M = 1.71%) using this measure. Consequently, a more sensitive measure would be required to capture accuracy developments of the targeted forms for the main study. In addition, the lack of gains in accuracy as well as complexity from the unguided planners was probably due to the fact that their attention was not drawn to the targeted forms during planning, and as a result, they completed the tasks using other linguistic structures.

In terms of fluency, there were mixed results for the B1 learners as the guided planners produced larger gains at the immediate post-test whilst the unguided planners produced larger gains at the delayed post-test. In terms of the B2 learners, it was the unguided planners who produced the largest gains in fluency. Thus, the results of this study show that tasks which incorporate guided planning and are sequenced according to an increase in cognitive complexity can facilitate L2 oral development in terms of accurate use of RCs, cognitive state verbs, as well as gains in syntactic complexity but not necessarily fluency compared to unguided planning and task complexity for this sample of B1 and B2 learners.

So how can we account for the lack of gains in fluency from the guided planners? It appears that the guided planning conditions towards the targeted forms resulted in the GP learners attending to the linguistic demands of the tasks at the post-

tests. As we saw in 3.3.2, Robinson (2011) argues that complex tasks draw learners' attention and effort to encoding the linguistic demands of a task that results in gains in accuracy and complexity at the expense of fluency. This appeared to be the case at the post-tests for the B2 guided planners where they may have prioritized accuracy over fluency as shown in the pre- immediate post-test gain for accuracy (*M* = 18.57) using 'error-free relative clauses per relative clause', and the drop in fluency (*M* = -18.35) using the fluency 'speech rate a' measure. The B2 unguided planners however, were not drawn towards form during their planning treatment and therefore may have focused more on meaning and the storyline at the post-tests which appears to have benefitted fluency, for example, the pre- immediate gain (*M* = 13.48) at the expense of accuracy (*M* = -50) using the same measures.

In addition, these results showed that intermediate level learners appear to have more stable declarative knowledge of relative clauses compared to the beginner learners in Mochizuki & Ortega (2008) and thus were able to capitalize on the gains afforded by guided planning and task complexity which had positive consequences for L2 relativization development. However, the treatment did not result in over-use of the targeted forms as there were only two instances during the pre- and post-tests where two participants (B1UP and B2GP) produced more than seven of the expected RCs which occurred during the delayed post-test. Thus, over-use of the forms was not an issue in the pilot study.

In relation to research question two, the B1 and B2 learners who received guided planning generally focused on the guidance provided in this case, relative clauses and cognitive state verbs, and this strategy remained largely unchanged as they progressed with more complex tasks. Likewise, the B1 and B2 learners who received unguided planning also focused on form, despite receiving no instruction to do so, and they maintained attention towards form as they progressed with more complex tasks. Thus, Japanese intermediate learners appear to have a general tendency to focus on form when planning. This sample did not appear concerned with planning for meaning in order to improve fluency even when given unguided planning conditions. Furthermore, they did not seem to attend to different aspects of L2 speech as they progressed with complex tasks over time. Instead they seemed to prioritize form over meaning. Clearly, guided planning conditions orientate Japanese learners towards form, but in the case of unguided planning, their focus on form could lye in their previous educational contexts which we discussed in 3.2.5 were heavily rooted in form-focused grammatical

instruction. Given these results, it appears that Japanese intermediate learners have a tendency to focus on form when planning for oral tasks which remain unchanged as they progress with more complex tasks over time.

4.7 Chapter conclusion

As noted at the start of this chapter, a pilot study is intended to identify any potential issues that may hinder the design of a main study. On the whole, the pilot study was considered a success in terms of answering research questions one and two. In terms of research question one, guided planning and task complexity successfully resulted in L2 oral development in terms of accuracy, and syntactic complexity compared to unguided planning and task complexity but not necessarily in terms of fluency. Regarding research question two, we were able to see that this sample of B1 and B2 level Japanese learners who received either guided or unguided planning generally attended to form as they prepared for tasks that increased in complexity over time. These results helped generate new ideas to improve the quality of the design for the main study. A number of amendments were subsequently made for the main study and they will be addressed in the following chapter. Below is a summary of them:

1. In order to carry out inferential statistical analysis using SPSS software, a larger sample of participants was needed. Previous task-based studies (Kawauchi, 2005) used a minimum of 11 students per group. In order to create a larger sample the main study focused on one intermediate level. As the results of the pilot study generally showed little difference between the B1 and B2 learners, it was decided that the main study would focus on B2 learners only, with approximately 15 learners per group. In addition, to create a larger group sample, the main study did not use a control group. As the results of the pilot study clearly showed no effect from the control groups in producing the targeted forms, they were not required for the main study. Rather, comparisons were made between two larger but different planning groups.
2. The pilot study investigated simple RC types (OS, OO), however given the improvements made in the use of the forms by the guided planners, a more

difficult RC type was included in the main study: 'OPREP', for example, *'the man likes the dog which the boy is looking at'.*

3. Due to the B1 and B2 guided planners' success in producing RCs at the post-tests, it was decided that a greater linguistic challenge would be to test learners' use of the correct verb tense within relative clauses by designing narrative tasks that contain singular and plural use of the head noun located next to the RC, for example, 'He thinks he likes the *dogs which have* long hair'vs 'He thinks he likes the *dog which has* long hair.' This involves correct use of 3rd person plural vs 3rd person singular respectively. Consequently, the main study focused on RCs and the accompanying use of 3rd person singular and 3rd person plural, as opposed to cognitive state verbs.
4. Previous task-based acquisition studies (Mackey, 1999, Baralt, 2010) have used productive and receptive tests to confirm acquisition of grammatical forms. In order to test development of RCs in the main study, a receptive test was designed and used. The narrative tests remained given their success in eliciting the targeted forms.
5. As the results of this study showed that the guided planners clearly produced more accurate instances of the targeted forms compared to the unguided planners, the unguided planning condition of the pilot study was replaced by an alternative planning condition that could provide more pedagogically useful results in terms of the development of the targeted forms. Clearly, in order for a group of learners to produce the targeted forms, their attention must be directed towards them in some way. It was therefore decided that a group of learners would receive initial guidance towards the targeted forms, but then be left to plan independently on subsequent task attempts, referred to as 'guided and unguided planning'.
6. The holistic nature of the accuracy 'error-free' measures used was considered unsuitable as a sensitive measure of L2 development. Consequently, it was decided that the main study would use an alternative measure for accuracy involving Mochizuki & Ortega's (2008) rating scale which was a more sensitive measure for tracking grammatical developments of the targeted forms.
7. In terms of research question two, in order to reduce the subjectivity of the qualitative analysis concerning the learners' planning strategies, the main

study did not include hypotheses. In doing so, the interview data could be interpretated without any preconceived perceptions, and as a result, the findings of the study would be considered less impartial.

8. Finally, to reduce the subjectivity of the qualitative data analysis even further, the main study used computer assisted qualitative software that could perform word frequency counts. In addition, a pre-treatment questionnaire was also used in order to find out additional information regarding the participants' perceptions towards L2 speaking before engaging in their respective planning conditions.

In light of these revisions, as well as taking into consideration the results of the pilot study, our original research questions and hypotheses were modified slightly for the main study as follows:

- Research question one: to what extent does guided planning and task complexity facilitate L2 oral development in terms of fluency, morphological accuracy involving OS and OPREP RC types and 3rd person singular and plural, as well as syntactic complexity for second year Japanese university learners of English?

- Hypothesis one: guided planning and task complexity will facilitate L2 oral development to a lesser extent than guided and unguided planning and task complexity in terms of fluency.

- Hypothesis two: guided planning and task complexity will facilitate L2 oral development to a greater extent than guided and unguided planning and task complexity in terms of morphological accuracy of OS and OPREP RC types and 3rd person singular and plural.

- Hypothesis three: guided planning and task complexity will facilitate L2 oral development to a greater extent than guided and unguided planning and task complexity in terms of syntactic complexity.

- Hypothesis four: guided planning and task complexity will facilitate L2 development to a greater extent than guided and unguided planning and task complexity in terms of learners' receptive awareness of OS and OPREP RC types and 3rd person singular and plural.

- Research question two: what strategies do Japanese second year university learners of English use when planning for oral narratives that increase in complexity over time?

5. METHODOLOGY

5.1 Introduction

This chapter concerns the methodology for the main study. As we saw in the previous chapter, the results of the pilot study led to a number of amendments for the main study and also resulted in the modification of our research questions and hypotheses. This chapter will therefore discuss all the changes that were implemented into the main study in order to answer research questions one and two. As we have already discussed the merits of quantitative, qualitative and longitudinal mixed methods research for the design of this study in 4.2 and 4.2.1, we now move onto discuss the components of the main study. This chapter begins with an outline of the study design followed by a description of the participants and the targeted linguistic forms. The pre- and post-tests are then discussed in 5.3, the treatment tasks and planning conditions in 5.4. The procedure is outlined in 5.5. In 5.6 we discuss the post-task interviews as well as justifying the use of a questionnaire. Finally, 5.7 describes and justifies the data analysis of the main study.

5.2 Experimental Design

Following on from the design of the pilot study in 4.3 and previous developmental studies that have incorporated pre- and post-tests (Bygate, 2001b; Mackey, 1999) this study employed a mixed factorial repeated-measures design. In 4.3 we saw that repeated-measures designs are commonly used within applied linguistics research and are relevant for quantitative longitudinal studies that involve analysing a sample of participants over certain points in time (Dornyei, 2007). This study consisted of a mixed factorial design that incorporated within-subject and between-subject factors. The within-subject variable was testing with three levels: pre-test, immediate post-test, delayed post-test. An additional feature of the study was task complexity as all the students performed a sequence of tasks that increased in complexity along intentional reasoning demands. Finally, the between-subject variable was planning time with two levels: guided planning, and guided and unguided planning.

5.2.1 The participants

Thirty four Japanese students of English from Shimonoseki City University in Japan participated in the study. They were all aged between eighteen to twenty one years old, with a mean age of nineteen. Eight students were male and twenty six were female. The participants had studied English for approximately the same length of time: seven to nine years, and some of them had travelled abroad as part of study-abroad programs. Twenty seven participants were second year university students majoring in Economics or International Commerce and were enrolled in an English language oral communication program as an additional credit to their course. The remaining seven participants were third year students. Six of them had already graduated from the English language course and one student was repeating the program. The seven third year students were primarily recruited to counter the threat of attrition outlined in 4.2.1 and were included in case any of the second year learners withdrew from the study.

Unlike the students who participated in the pilot study from Ritsumeikan Asia Pacific University, most of the participants in the main study did not have TOEIC or TOEFL scores. They were placed into different English level classes based on the results of an internal English language placement test. As we saw at the end of the last chapter in 4.7, the intention of the main study was to recruit B2 oral level learners in order to match the proficiency of the learners who participated in the pilot study. However, as the placement test at Shimonoseki City University contained no speaking component, it was decided that students would be recruited from the intermediate level classes and above. Prior to the study, each student met with the researcher individually to discuss the details of the study in English. This led the researcher to confirm that the sample as a whole represented similar L2 oral ability to the B2 learners from the pilot study. As pointed out in 4.3.1, the CEFR's (2010) definition of B2 oral proficiency is that learners "can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party" (p. 24). This descriptor appeared to be an appropriate match of the L2 oral ability of the overall sample recruited for the study.

In 4.7 we mentioned that the intention of the study was to compare two experimental groups. The thirty four participants were therefore assigned into two groups: guided planning (GP), and guided and unguided planning (GUP). Both groups

were considered relatively homogenous in terms of n-size (n = 17), gender, and background. However, as all of the 27 second year students completed the study, the seven third year learners were later removed from the analysis as they did not participate in the placement test. Although removing these seven learners would reduce the sample size it would increase the probability that both groups were evenly matched as we could now compare both groups' mean placement test scores to ensure that they were considered equal in terms of proficiency. This was achieved by carrying out an independent samples *t*-test to compare the means of both groups' placement test scores to see whether there were any significant differences between them. A *t*-test was chosen because it is a common form of analysis when measuring statistical significance between the means of two groups (Dornyei, 2007). A *t*-test reports a probability (*p)* value, which if equal to or less than .05 ($p < 0.05$) implies there was a significant difference between the means of the two groups (Field, 2009), and that the results could only have occurred by chance in 0.05% of cases. First, although the seven third year students were initially placed relatively evenly within each group (four in the GP group and three in the GUP group), the removal of them required one second year student to also be removed in order to ensure an equal n-size per group. Thus one male second year student from the GUP group was also removed so that both groups contained thirteen learners and were still relatively homogenous in terms of n-size (n = 13) gender (the GP group now contained three males and ten females whilst the GUP group contained two males and eleven females), age, and background. The results of the *t*-test showed that the GP group averaged a slightly larger placement test score of (M = 561.38) compared to the GUP score of (M = 544.62) but this difference was not significant $t(24) = 1.262$, $p > 0.05$. As a result, both groups were also considered homogenous in terms of proficiency. Furthermore, as we will see in 6.2 when we analyze both groups' oral performances, further *t*-test results show there were no significant differences between both groups' pre-test scores thus confirming homogeneity in terms of L2 oral ability prior to the study. Thus, the group sample size of the main study is displayed in table 20:

Table 20: *The groups*

Guided planners (GP)	13 learners
Guided and unguided planners (GUP)	13 learners

However, when we go on to discuss the data analysis in 6.2, we will see that it was necessary to reduce the group sample size to (n = 12) for two measures related to fluency and complexity in order to ensure both groups could be considered equal in relation to their pre-test performance.

The purpose of comparing two groups of B2 oral Japanese intermediate learners of English were two fold, firstly the findings of this study would allow us to draw comparisons with the majority of previous planning studies which also used intermediate level learners (Ellis, 2009a). Secondly, we know the results of Mochizuki & Ortega (2008) were disappointing in terms of the overall sample's production of RCs. One of the contributing factors for this was the participants'' beginner level proficiency. However, we saw from the results of the pilot study that Japanese intermediate learners appear to benefit from the treatment of guided planning and task complexity with clear developmental gains in accuracy of the targeted forms thus warranting further investigation on a larger scale. The sample size of thirteen students per group is approximately in line with Dornyei's (2007) recommendation of using fifteen learners per group for experimental studies. Field (2009) suggests using a minimum number of thirty participants in order to perform quantitative statistical analysis and although the total sample of this study was thirty four, several participants were removed in order to ensure the comparison of two equal groups as discussed above.

Further justifications of our group sample sizes can be made by comparing them with previous planning studies in which the lowest group sizes appear to be Kawauchi (2005) who compared task performance between eleven advanced learners, twelve upper-intermediate learners and sixteen low-intermediate learners. Although other previous planning studies have used larger samples, all of them except of Bygate (2001b) were cross-sectional studies that only required students' participation at one point in time. As we discussed in 4.2.1, it is more difficult to recruit large sample sizes for longitudinal designs because they require more time commitment from learners, and as a result, they may withdraw from the study. However, the financial incentive

stipulated in the participants' consent form (see appendix J) prevented this from occurring, and on completion of the data collection, the researcher was content with an n-size of 13 per group in order to carry out inferential statistical analysis.

Although we can justify our sample size, it is nevertheless a small one, and as a result, it is important not to over-generalize the results. According to Dornyei (2007) "a good sample is very similar to the target population in its most important general characteristics (for example, age, gender, ethnicity, educational background, academic capability, social class or socioeconomic status)" (p. 96). In terms of the present study, our target population refers to Japanese second year intermediate university learners of English of which our sample appears to be an accurate representation as discussed above. Consequently, providing we suggest that the findings of this study provide indications about the target population then we can argue that the findings of this study can be generalised to similar populations of Japanese B2 intermediate university level learners.

5.2.2 Main L2 target forms: OS and OPREP RC types

The main linguistic features targeted for the present study was the simple RC type 'object-subject' (OS) that was used in the pilot study, for example, 'she wants the doll *which has black shoes',* and the introduction of a difficult RC type 'object of a preposition' (OPREP), for example, 'he likes the dog *which the man is looking at.* These targeted forms were chosen for three main reasons. First, we know from 3.5.4 that English RCs are considered to be a difficult linguistic feature for Japanese learners to produce and acquire given the differences that exist between English and Japanese versions of the form. Although RCs are instructed to Japanese learners during the second year of junior high school, RC studies such as Schachter (1974) have reported Japanese intermediate learners' avoidance of the form in oral production. Second, we saw in 3.3.6 that the only previous study which examined Japanese learners' use of RC types in natural language use was Mochizuki & Ortega (2008) who investigated simple RC types including OS and the results of the study showed that the beginner level participants did not produce as many instances of the forms as expected. Consequently, the purpose of the present study was to see whether Japanese B2 intermediate learners would have more success in producing the same RC type. The results of our pilot study

in 4.5.1 showed clear gains in the B2 learners' error-free production of the OS form, and so the purpose of the main study was to pursue the development of the same RC type with a larger sample. Finally, given the higher proficiency of the B2 learners compared to Mochizuki & Ortega's (2008) study, one of the simple RC types used in the pilot study (OO) (she wants the doll *which the girl is watching*) was replaced with the more difficult RC type OPREP in order to provide a greater linguistic challenge for intermediate level learners. As mentioned in 3.5.3, Keenan & Comrie's (1974) NPAH hypothesis placed the 'object of a preposition' (OPREP) as being more difficult to acquire than simple RC types such as those used in Mochizuki & Ortega (2008). In addition, we saw in Erkman et al.'s (1988) study in 3.5.5 that OPREP instruction not only benefited OPREP production but it also provided generalized learning effects across less marked RC types such as OS. As a result, it was intended that instruction of the OPREP RC, through guided planning, would not only develop learners' use of OPREP production but may also assist in the development of OS use as well. The intention of the main study therefore was to use guided planning that focused on the simple OS RC type and the complex OPREP. In order to do this however, the narrative tasks would need to be adapted in order to facilitate use of the OPREP RC type. Furthermore, the guided planning conditions would need to illustrate OPREP examples in order to draw learners' attention towards using the form. 5.3.1 and 5.4.1 will demonstrate how this was achieved.

In addition to the targeted RC types, this study was also concerned with the grammatical structures that compliment RCs. The next sub-section shall examine this in more detail.

5.2.3 Relative clause relevant morphology: 3rd person singular and plural

"Morphology is concerned with the structure of words and phrases" (Carter & McCarthy, 2006, p. 2). For example, the noun *dog* is singular but adding plural 's' changes the meaning to the plural form *dogs*. Small linguistic changes such as plural 's' are referred to as grammatical morphemes which come under the umbrella of morphology. Relative clause relevant morphology concerns forms within the structure of RCs such as verb tense, for example, 'the dog which *has* long hair'. One way to

examine learners' oral development of RC types with correct verb tense would be to design narratives that facilitate instances of singular and plural use of the head noun, for example, 'he thinks he likes the dogs *which have* long hair' vs 'he thinks he likes the dog *which has* long hair'. In the former case, the morphology that compliments targetlike use of the OS RC type is 3rd person plural as the RC type is located next to the head noun 'dogs', whilst in the latter case, the relevant morphology is 3rd person singular as the RC type is located next to the singular head noun 'dog'.

3rd person singular also relates to how cognitive state verbs were used in the pilot study, for example, '*he thinks* that *he likes*….', 's*he believes* that *she likes*...'. As we saw in 3.4.3, cognitive state verbs were chosen in response to Robinson's (2007) study which reported their use when explaining the intentions of other people, for example, '*he thinks that…., he wonders that…'* and how the forms were compatible with RCs, for example, '*he thinks* that he likes the dog *which has long ears'*. However, as the pilot study only focused on specific cognitive verbs i.e. *'thinks', 'believes', 'likes'* and *'wants'* it was considered more appropriate that the main study should focus on the development of 3rd person singular as opposed to cognitive state verbs. Consequently then, the linguistic forms of the main study involve the OS and OPREP RC types and the morphological adequacy that accompanies them: specifically the use of 3rd person singular and 3rd person plural. Together, these linguistic forms allow us to analyse learners' development of intentional reasoning speech which requires learners to explain the actions of other people.

5.3 Pre- and post-tests

In order to measure L2 development, two different types of assessment were used: an oral narrative and a grammatical judgement test. The narrative was a continuation of the assessment used to test oral development in the pilot study in terms of fluency, complexity and accuracy. However, amendments were required in order to test learners' use of 3rd person singular and plural which will be described in the following sub-section. The second assessment was a grammatical judgement test which was used as an alternative assessment to test learners' receptive awareness of the targeted forms. This test served to compliment the narrative assessment which targeted learners' production of the targeted forms. These measures follow the view of Schmitt (2010) and Baralt

(2010), who argue that alternative assessments provide greater indication that L2 development of linguistic forms has taken place. In relation to task planning studies, Ellis (2009a) points out that no planning study has yet been able to show how the benefits of task planning (strategic planning or task repetition) can be transferred across to a new or different type of task. As a result, an important aspect of this study's testing procedures was to have participants perform different types of tests to see whether the effects of guided planning and task complexity can be transferred to other types of performance.

In addition, Ellis (2008) notes that it is important for developmental studies to try and identify the type of L2 knowledge that results from the treatment provided:

> SLA researchers do not always take note of the distinction between implicit and explicit knowledge, opting instead to simply investigate undifferentiated L2 knowledge and to talk about what learners 'have learnt' or 'know' without bothering about the nature of knowledge they are investigating. (p. 427)

Douglas (2001) notes that in order to identify the type of L2 knowledge acquired, researchers need to design constructs that could measure and confirm the type of learning that occurred:

> construct validity may be demonstrated by the construction of theoretical arguments linking hypothesized aspects of language ability to features of the test tasks, demonstrating the appropriacy of the tasks for making interpretations regarding the construct, and then providing empirical evidence that the links are in fact present. (p. 447)

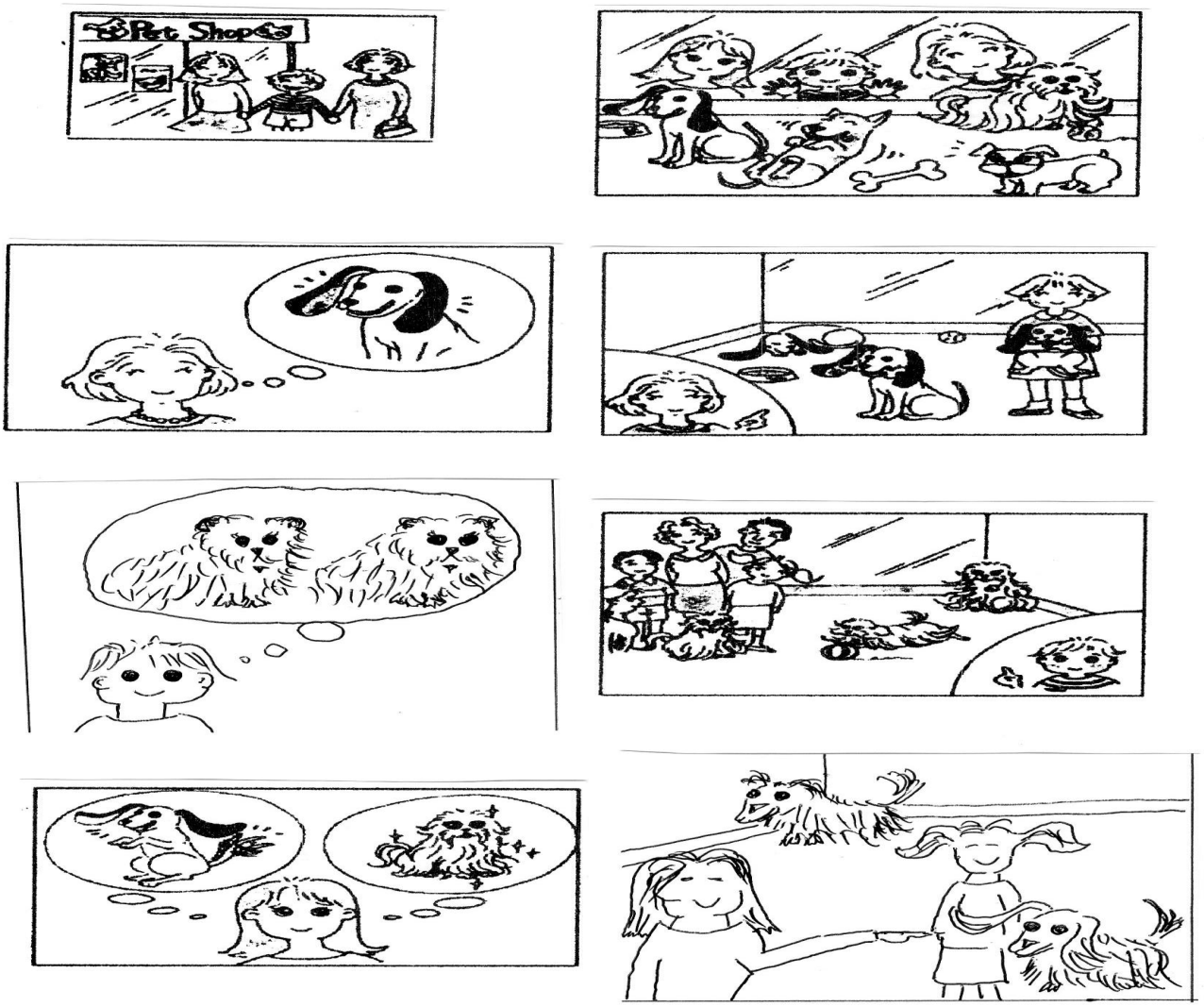
As we will see in the following two sub-sections, the present study took steps to ensure that both tests used to measure L2 development were designed to measure implicit knowledge by creating conditions for unplanned language use. We shall now describe and justify each assessment used.

5.3.1 The oral narrative

The first assessment was an oral narrative. The pre- and post-test narratives were adapted from the pre- and post-test narratives of the pilot study, which in turn, were based on the task used Mochizuki & Ortega (2008). Although slight changes were made

to accommodate use of the OPREP RC type as well as 3rd person singular and plural. Figure 9 displays the pre-test narrative.

Figure 9. *Pre-test narrative (adapted from Mochizuki & Ortega, 2008)*



The pre-test narrative storyline involved a boy called Kevin, his mother and sister who go to a pet shop. In line with Mochizuki & Ortega's (2008) focused task, the narrative consisted of eight pictures sequenced in correct order and it was designed to elicit 7 obligatory cases of RCs. For example, in one picture 'Kevin's mother *thinks that she likes* a dog *which has long ears'.* In addition, the task was designed to facilitate use of 3rd person singular, for example, certain pictures contained thought bubbles from the character's head illustrating their feelings which could elicit language such as *'*he *thinks* that he *likes* a dog*…'* However, the narrative needed adapting to facilitate use of 3rd person plural, as the narrative in Mochizuki & Ortega (2008) contained seven cases of

RCs in which the head noun was singular. Consequently, one picture was replaced by a new picture that attempted to elicit 3rd person plural. This picture represents RC context three and was intended to elicit language such as 'Kevin likes the *dogs which have long hai*r' (see figure 10 below). The second amendment to the task involved facilitating the use of the OPREP RC type. However, the original Mochizuki & Ortega (2008) narrative contained one picture (context four) that could elicit OPREP as it could be interpreted as 'Kevin wants the dog *which the family is looking at*.' Consequently, only one other picture was designed to elicit the OPREP RC (context seven) and it was intended to facilitate language such as 'Kevin's sister wants the dog *which the girl is looking at.*' This picture replaced the corresponding picture of the original task. The two new pictures however did not alter the main storyline. As a result, the pre-test could now facilitate seven cases of RCs that included both singular and plural head nouns including two instances of the OPREP RC. Possible examples for each RC context are displayed in figure 10:

Figure 10. *Pre-test narrative: possible examples for each relative clause context*

Context 1 (OS): The mother thinks that she likes the dog which has long ears
Context 2 (OS): She wants the dog which is next to the girl
Context 3 (OS): Kevin thinks he likes the dogs which have long hair
Content 4 (OPREP): He wants the dog which the family is looking at
Context 5 (OS): Kate thinks she likes the dog which has long hair
Context 6 (OS): She also thinks she likes the dog which has long ears
Context 7 (OPREP): She wants the dog which the girl is smiling at

The testing conditions remained the same as in the pilot study. No planning time was allocated prior to the test so as to reflect normal conditions for language use (Robinson, 2005). As we discussed in 3.5.2, testing conditions that involve no planning time allow us to test for implicit knowledge which is determined by fluent, unplanned language use (Ellis, 2008). Previous studies such as Ellis (2005b) (cited in Ellis, 2008) tested implicit knowledge by using an oral narrative that did not involve planning so as to prevent learners from consciously thinking about their speech prior to performance. The present study adopted a similar stance by following the no-planning conditions of Yuan & Ellis (2003) in which the participants were allowed to briefly look at the

narrative for approximately twenty seconds prior to performance which restricted their opportunity to plan. As a result, the narrative was considered complex along *resource-dispersing* dimensions because the participants had no time to prepare prior to performance.

Each participant was instructed to narrate the story in English using as much detail as they could to the researcher who acted as the listener. The researcher had a copy of the narrative but with the pictures mixed up and he informed each participant that he would listen to their narration and match the pictures accordingly. This was in accordance with Mochizuki & Ortega (2008) who also used this strategy to create a communicative incentive for the speaker. In addition, in line with Yuan & Ellis (2003), each participant was asked to begin the narrative by saying, 'Today, Kevin, his sister and Mother are at….' so as to provide conformity and use of the present tense.

Following the pilot study, three different versions of the oral narrative were designed: one for the pre-test, the immediate post-test, and the delayed post-test. All versions were designed to be the same in terms of cognitive complexity. As in the case of the pilot study, each test version was similar in terms of resource-directing dimensions, in other words, the linguistic demands of each task were the same. For example, each oral narrative contained seven cases of RCs (five OS RC types with context three eliciting plural use of the head noun, and contexts four and seven eliciting two OPREP RC types) as well as the same instances of 3rd person singular. For example, each narrative contained three pictures of different people thinking about something. Although the pre- and post-test narratives were the same in terms of resource-directing dimensions, they differed in terms of storyline and characters for example, the immediate post-test narrative involved Tim, his mother and father going to a clothes shop, whilst the delayed narrative test involved three sisters going to a garden centre (see appendix K).

The advantage of using an oral narrative to test RC development is that it enables learners to use the form in natural language use. According to Ellis (2008), previous RC acquisition studies are limited in the sense that they did not use oral tasks, opting instead to use controlled oral exercises such as sentence combination tests as in Doughty (1991) and Izumi (2003). The use of a narrative in the present study allows for comparisons to be made with the results of the pilot study as well as the findings of Mochizuki & Ortega (2008) which also used narratives. In terms of disadvantages, we saw in 3.2.4 that using narratives to elicit targeted forms can result in learners avoiding

them whilst still completing the task (Seedhouse, 2005). Furthermore, as mentioned in 3.2.1, Loschly & Bley-Vroman (1993) argue that tasks should be designed to make targeted forms as essential as possible for L2 production. However, the results of Mochizuki & Ortega (2008) showed that this is difficult to achieve as their study displayed low values of RC production despite efforts in task design to elicit the form. Despite these disadvantages, the present study continued with a narrative as a test for development, largely based on the results of the pilot study which showed clear accuracy gains in the development of the forms for the B2 guided planners. The results showed that guided planning and task complexity facilitated the proceduralisation of the forms which enabled the B2 learners to produce them during the post-test narratives. Thus the narrative was considered a suitable oral assessment of the targeted forms.

5.3.2 The grammatical judgement test

The second assessment was a grammatical judgement test that was adapted from Izumi (2003) which investigated learners' acquisition of 'subject', 'direct object' and 'object of a preposition' RC types. The test consisted of 49 items (see appendix L). In accordance with the grammatical judgement test in Izumi (2003), 36 of the items related to the three RC types. 12 sentences represented the 'subject' RC type, for example, *'I like the dog which has long ears'*. 12 sentences represented the 'direct object' RC type, for example, *'You met the woman who went to the hospital'*, and another 12 sentences represented the 'object of a preposition' RC type, for example, *'I like the dog which the people are looking at'*. Although two of the RC types 'subject' and 'object of a preposition' were the targeted RC types in the present study (OS and OPREP), the 'direct object' RC type was not, therefore its inclusion is a limitation of this test. Nevertheless, the 'direct object' is classified as a simple RC type, in line with the 'subject' RC, therefore it was of interest to see whether guided planning towards the OS and OPREP RC types benefited the development of the 'direct object' RC type as well.

In accordance with Izumi (2003), the sentences of each RC type involved six correct items and six incorrect items. The incorrect items were based on four common errors outlined in previous RC studies such as Doughty (1991), for example, "pronoun retention, nonadjacency, incorrect relative marker morphology, and inappropriate

relative marker omission" (Izumi, 2003, p. 300). Izumi (2003, p. 300) illustrates examples of these errors as follows:

- The woman who you met her went to the hospital (pronoun retention)
- The woman is young who likes John (nonadjacency)
- I looked for the book who Tom was talking about (incorrect relative marker morphology)
- The girl was in pain saw the dentist (inappropriate relative marker omission)

For the purpose of this study, an additional error-type was included that targeted incorrect RC verb tense in relation to the head noun, in order to test learners' use of 3rd person singular and plural. For example:

- He wants the car which have big wheels

In line with Izumi (2003), the six correct items for each RC type, three of them contained the RC inserted in the subject position and three items contained the RC inserted in the object position. This was also the case for the six incorrect items for each RC type. In addition, the test used in this study was further adapted to test learners' receptive knowledge of 3rd person singular when accompanying RCs. Six items represented the form, three were correct and three were incorrect. The three correct 3rd person singular items included OS RC types in which the RC was inserted in the object position. For example, *'she thinks that she likes the man who has long hair'.* In doing so, the participants could be tested on the correct use of similar phrases used during their guided planning treatment involving 3rd person singular and the accompanying OS or OPREP RC types. The three incorrect items of 3rd person singular involved incorrect subject-verb agreement that accompanied an RC which was inserted in the object position so that the participants were again tested on similar phrases used during guided planning. For example, *'Peter believe that he played a piano which was made in France'*. Thus the incorrect 3rd person singular items were based on one error:

- incorrect subject-verb agreement

Finally, in an attempt to prevent learners from identifying the target forms of the study, 10 distracter items were included in the test that did not contain RCs. The distracter items were taken from Reinders & Cho (2012) which also used a grammatical judgement test. An example of a distracter was "he completed his study successful" (p. 25). Although no instructions were provided regarding the targeted RC types, the grammatical judgement test was considered to be the most likely test to indicate the study's target forms. Consequently, in line with Baralt (2010), this test was administered after the narrative to minimize the chance for the learners to identify the target forms. Finally, in line with Izumi and Izumi (2004), to prevent a learning effect from using the same test during the pre- and post-tests, the items of the grammaticality judgement test were rearranged for the immediate post-test and the delayed post-test, and random lexical items were exchanged with equally straightforward words.

The benefits of using grammatical judgement tests for developmental studies are that they allow researchers to construct sentences that contain the target feature under investigation (Murphy, 1997). Researchers can then construct correct and incorrect instances of the target feature in order to test learners' knowledge of the form. This was the case in the present study where we were able to design items that specifically targeted OS and OPREP RC types as well as correct and incorrect use of 3rd person singular and plural. In addition, grammatical judgement tests have been widely used in the literature as an assessment tool therefore the present study was able to refer to previous studies such as Izumi (2003) which was a useful source for designing grammatical judgement tests for RCs. Indeed Robson (2002) points to the advantage of using similar tests because it enables comparisons to be made with other studies.

On the other hand, grammatical judgements tests have been criticised for their failure to show why a learner may have judged an item to be grammaticality incorrect (Sorace, 1996). In order words, a learner could simply be guessing that an item is incorrect thus making their response invalid or the learner may think an item is incorrect based on a grammatical structure that was not the intended target structure of the study. To avoid these issues, Izumi (2003) and Doughty (1991), designed their tests to ensure that learners were judging an item to be incorrect based on the targeted RC types. This was achieved through learner error correction. For example, if a participant considered an item to be incorrect, they were required to write the correct version in a space provided. The participant would receive a mark only if their correction related to the targeted RC type. The present study therefore applied the same scoring system as Izumi

(2003). The learners were handed their test paper sheet and the researcher provided the instructions orally. The students were not informed about the target forms, and were simply told that they were taking a grammar test. They were instructed to mark with a tick if they thought the sentence was correct, and to mark with a cross for any incorrect sentences and to then write the complete correct sentence in the space provided.

In Izumi's (2003) study, learners would listen to an aural recording of each item and were then given 15 seconds to answer. The present study did not use an aural recording for each item therefore an estimated 20 seconds was allocated for the students to read and answer each item which set an overall time limit of 17 minutes to complete all 49 items. The time limit also enabled the test to serve as an indicator of implicit knowledge as well as receptive knowledge. As discussed in 3.5.2, implicit knowledge can be determined in unplanned language use. Previous studies such as Ellis (2005b) (cited in Ellis, 2008) tested implicit knowledge by using a timed grammatical judgement test to prevent learners from consciously thinking and planning their answers. Thus, as both assessments of the present study limited learners from engaging in conscious planning they were used as indicators of implicit knowledge. However, as the narratives involved a number of changes from the pilot study, and the grammatical judgement test was new, it was important to pilot the assessments to ensure the adapted targeted forms were a suitable linguistic challenge for B2 oral learners. The next sub-section discusses this in detail.

5.3.3 Piloting the pre- post-tests

The grammatical judgement test was piloted on ten Japanese intermediate students who did not take part in the study but belonged in the same class as the participants who did. As this class was a higher level in terms of the students' average placement test score, it was expected that their performance in the test would resemble a slightly higher average than the overall sample of the main study which incorporated learners from lower-level classes as well. The test consisted of 40 items which were randomly selected from the pre- and post-tests. The items consisted of correct and incorrect examples of the targeted RC types (OS and OPREP) including RCs that were located next to singular and plural head nouns, as well as items that contained 3rd person singular with correct and

incorrect subject-verb agreement. A small number of distracter items were also included. A time limit was set for ten minutes.

The results of the grammatical judgement test showed that most students scored incorrectly with RCs that had singular and plural head nouns, as well as errors in connection with 3rd person singular. In addition, there were a small number of errors relating to the OPREP RC type. The average score of the test was 67.8%, and as this sample consisted of learners who were placed in the higher level English classes as a result of their placement test, the targeted forms were considered to be a suitable linguistic challenge for the main study participants.

In addition to the grammatical judgement test, the pre-test narrative was piloted two Japanese university students of English to see whether they could orally produce the adapted targeted forms of the main study. Both of these students did not participate in the main study but were considered to be of similar B2 oral level proficiency. Similar guided planning conditions were provided to the main study in which the researcher first provided a ten minute workshop explaining the guided planning conditions, each student was then allocated ten minutes guided planning time before narrating each task. The results of their performance showed that both students produced instances of the targeted RC types including the OPREP RC, as well as 3rd person singular. Based on these findings the tasks were expected to elicit the adapted targeted forms.

5.4 The treatment tasks

The tasks designed for the main study were adapted from the treatment narratives used in the pilot study. As we know from 4.3.4, the pilot narratives were based on the focused task used in Mochizuki & Ortega (2008) which was designed to elicit seven obligatory contexts of simple RC types. However, the main study was interested in eliciting the complex RC type OPREP and use of 3rd person singular and plural so the pilot narratives were adapted to accommodate these changes. Five narratives were designed by the researcher for the treatment sessions (see appendix M). Each task contained a different story line following Samuda & Bygate's (2008) recommendation that sequencing tasks with different storylines helps to maintain learners' interest as opposed to repeating the same task. Consequently, the same narrative storylines from the pilot study were used for the narratives in the main study as they involved different

characters and locations, for example, a pet shop, a toy store and a car shop. In line with the pilot study, the treatment tasks were also different in terms of cognitive complexity. The treatment narratives were designed and sequenced to increase in complexity according to the claims of Robinson's (2003) Cognition Hypothesis which was outlined in 3.4.2. This involved sequencing the tasks so they increased in complexity along resource-directing dimensions by increasing the intentional reasoning demands of them. This was achieved by designing pictures that contained additional instances of characters thinking about something, and then choosing something so as to elicit RCs and the accompanying use of 3rd person singular or plural. For example, narrative one contained seven RCs, narratives two and three contained nine instances whilst narratives four and five contained ten instances of RCs (see table 21).

Table 21. *Obligatory cases of RCs per narrative*

Targeted Forms	Pre-test Narrative	Task treatment (increasing intentional reasoning demands)			Immediate post-test narrative	Delayed post-test narrative
		Narrative 1	Narrative 2 & 3	Narrative 4 & 5		
Obligatory cases of RCs	7	7	9	10	7	7

Having discussed the treatment tasks and how they were sequenced, we now move onto discuss the GP and GUP group's planning conditions for the tasks.

5.4.1 Guided and unguided planning conditions

Alterations were made to the planning conditions of the main study based off the results of the pilot study. As the B2 guided planners in the pilot study successfully developed accurate use of RCs, the researcher was content to continue with similar guided planning conditions for the main study. However, the guided planning notes were adapted to accommodate changes with the targeted RC types. They included four example phrases of the RC types (two 'OS' and two 'OPREP') and accompanying use

of third person singular. For example; *'Kate thinks that she likes the cat which has blue eyes'* (see appendix N).

As pointed out at the end of the last chapter in 4.6, alterations were made to the planning conditions of the unguided planning group on account of their disappointing results in terms of RC accuracy development. As we discussed in 4.7, these results were probably attributed to the learners' lack of attention towards the targeted forms as a result of the unguided planning conditions. It was therefore decided that the unguided planning conditions should provide some guidance towards the targeted forms. Consequently, the second group of learners (GUP) would receive initial guidance towards the forms but were then left to plan independently on subsequent task attempts. This condition was referred to as 'guided and unguided planning'. This condition would allow us to see whether learners of B2 oral proficiency, who already had explicit knowledge of the targeted forms, may simply require initial attention towards them, after which, they may be able to use their own linguistic resources to plan independently over time. The guided and unguided planning group therefore received the same guided planning treatment as the guided planners at the beginning of the treatment in week two. However, in weeks three and four, they received unguided planning in which they were allocated blank note-sheets and were verbally instructed to think about content and language in accordance with previous unguided planning conditions (for example, Yuan & Ellis, 2003). Finally, the control group of the pilot study that involved no planning time was removed from the main study due to the similar results of the unguided planning group that lacked output relating to the targeted forms. As a result, the main study involved the comparison of two different planning conditions: guided planning (GP), and guided and unguided planning (GUP).

In terms of the length of planning time, both groups were allocated ten minutes for the first treatment narrative in week two. Ten minutes was the same limit set in the pilot study, and was also the standard limit allocated to the majority of previous planning studies (Ellis, 2009a). Both groups received the grammar guidance notes and were instructed to try and use the grammar when they performed the task, and in accordance with Yuan & Ellis (2003), they were not allowed to use their notes when speaking. The task in week two was therefore simple along *resource-dispersing* dimensions because planning time was allowed. However, unlike the pilot study, as the treatment progressed, planning time for both groups gradually reduced in line with the claims of Robinson's (2003) Cognition Hypothesis outlined in 3.4.3 which states that

resource-dispersing dimensions should be increased during task sequencing in order to prime learners to perform tasks under the natural time constraints of everyday speech (Robinson, 2010). It was expected that gradual reduction in planning time during the treatment would help improve both groups' performance during the post-test narratives that involved no planning time. As a result, narratives two and three consisted of seven minutes planning time whilst narratives four and five consisted of four minutes (see table 22).

Table 22. *Strategic planning conditions during the treatment sequence*

	Narrative 1	Narrative 2 & 3	Narrative 4 & 5
GP	Guided planning	Guided planning	Guided planning
GUP	Guided planning	Unguided planning	Unguided planning

Note. Length of planning time = narrative 1 (10 minutes), narratives 2 & 3 (7 minutes), narratives 4 & 5 (4 minutes)

5.5 Procedure

Data collection for the study took place at Shimonoseki City University, Japan during June-July 2012. The study was carried out in a spare classroom and was not part of a course program. The duration of the main study remained at seven weeks in line with the pilot study (see table 23).

Table 23. *Study procedure*

Groups	Pre-test Week 1	Task complexity treatment (+ intentional reasoning demands) Week 2	Week 3	Week 4	Immediate Post-test Week 5	Delayed Post-test Week 7
GP	Narrative GJ Test	Training Session. Narrative 1 Post-interview	Narrative 2 & 3 Post-interview	Narrative 4 & 5 Post-interview	Narrative GJ Test	Narrative GJ Test
GUP	Narrative GJ Test	Training Session. Narrative 1 Post-interview	Narrative 2 & 3 Post-interview	Narrative 4 & 5 Post-interview	Narrative GJ Test	Narrative GJ Test

Note. GP = guided planning group, GUP = guided and unguided planning group, GJ = grammatical judgement

In week one, the pre-tests were carried out. Each student first performed the narrative with the researcher then the students performed the grammatical judgement test in small groups. In week two, all the participants first completed a questionnaire that investigated their perceptions towards communication and accuracy prior to engaging in the planning treatment (described in more detail in 5.6.3). Then in small groups, the learners took part in a 15 minute guidance session with the researcher which focused on correct use of the targeted forms: OS and OPREP RC types, and correct use of 3rd person singular and plural. The training session involved describing correct use of the forms using examples sentences followed by eliciting learners' production of the forms using some of the pictures from the pre-test narrative. Each participant then took turns performing treatment narrative one under the guided planning conditions outlined in the last sub-section. After the task, each student was interviewed one-on-one with the researcher regarding their planning strategies. In week three, the GP learners performed treatment narratives two and three in the same format as in week two except planning time was reduced from ten minutes to seven minutes. The GUP learners also received seven minutes planning time but they were not provided with the grammar guidance notes and were instructed to plan independently. After completing the tasks, the students were interviewed one-on-one with the researcher regarding their planning

strategies. During week four, both groups performed narratives four and five under similar conditions as week three. The GP learners received guided planning whilst the GUP planned independently. During this treatment session, both groups were allocated four minutes planning time. After the tasks, the participants completed the post-task interview. At week five, the GP and GUP learners performed the immediate post-tests under the same conditions as the pre-tests in week one. Finally, after a two week interval the learners participated in the delayed post-tests.

So far we have discussed the overall design of this study and the materials relevant for research question one which are the pre- and post-tests, the treatment tasks and the planning conditions. We now turn to the techniques needed to answer research question two which concerns learners' planning strategies. We first begin by discussing the advantages and disadvantages of post-task interviews, otherwise known as retrospective interviews or stimulated recall, followed by the merits of questionnaires.

5.6 Retrospective interviews / stimulated recall

According to Dornyei (2007), "in 'retrospection', the respondents verbalize their thoughts after they have performed a task or mental operation" (p. 148). As this study is investigating what learners think during strategic planning, retrospection appears to be an ideal technique for collecting data. One method that allows learners to comment on their thought processes is stimulated recall which is "used to prompt participants to recall thoughts they had while performing a task or participating in an event" (Gass & Mackey, 2000, p. 17). For the purpose of this study, stimulated recall would involve interviewing a participant after they had planned and performed an oral task so they could explain the strategies they used during strategic planning prior to performing a task. Dornyei (2007) points out that in order for learners to comment on previous thought processes, they need to retrieve the information from their long-term memory therefore the validity of the data collected would depend on the time lapse between task performance and the interview. The longer the time lapse, the more difficult it is for learners to retrieve accurate information regarding the event. In order to assist learners in accessing their thoughts, a stimulus is often used, for example, listening to a recording of their task performance, watching a video, or looking at written work which the learner produced.

Dornyei (2007) outlines the advantages of using stimulated recall which are summarized as follows:

1. It enables us to access the cognitive processes that trigger language production which cannot be accessed by other means.
2. Retrospective interviews can be carried out using a variety sources (videos, transcriptions, questionnaires).
3. It can be used with other research methods, it often provides rich data, and it can improve the reliability of the data analysis.
4. Critics of stimulated recall have never doubted people's ability to be able to retrieve their thought processes.
5. All major theoretical frameworks concerned with thinking have advocated the use of verbally reported sequences of thoughts.

A number of disadvantages in using stimulated recall are explained in Cohen et al. (2005) for example, points 1-3 below, whilst points 4 and 5 are reported in Dornyei (2007):

1. Verbal reports are susceptible to social desirability bias in that a participant may respond in a way that is deemed to be socially acceptable, thus masking their true feelings.
2. Participants' responses may also be influenced by background knowledge or social status, in other words, students' knowledge of the topic or their ethic background may bias their responses in terms of what they actually experienced.
3. The time lapse between the task performance and the interview may result in participants forgetting their thought processes. In this case, learners could forget the planning strategies they used prior to task performance.
4. Learners' awareness of producing verbal reports can have reactive consequences on their performance of the task and their thought processes. For example, learners may plan differently for a task if they are aware that they have to provide a verbal report and this may contaminate the data.
5. Learners may not be able to retrieve certain thought processes from their long-term memory because cognitive processing is often considered to be an

unconscious process whilst even conscious processes are often too complex to be re-told verbally.

These disadvantages, although valid in their own right, can be minimized to safeguard the validity of using stimulated recall for the purpose of this study. We will now discuss each point in turn.

1. In terms of social desirability bias, the researcher addressed this issue by informing the participants that their answers were kept strictly anonymous and their personal details were not included in the study. This was also outlined in the participants' consent form. Consequently, these measures were expected to reduce the possibility of biased responses.

2. Background knowledge bias was addressed by using the same approach as Ortega (1999) who avoided using leading questions so the participants were not aware of the purpose of the interview, and thus their responses were based on what they actually did. For example, avoidance of questions such as 'do you think planning helps improve your English?'

3. The issue of time lapse was nullified by following Dornyei's (2007) recommendation of keeping intervals between the interview and the task planning as short as possible, ideally 24 hours. In Ortega's (1999) study, retrospective interviews were carried out immediately after the task to cancel out a time lapse. This study followed suit by conducting interviews immediately after learners had planned and performed a task.

4. In dealing with reactive consequences of interviews, Dornyei (2007) recommends not informing the participants of the exact purpose of the interview before they participate in the task "so that the foreknowledge does not affect their performance" (p. 149). Ortega (1999) did not inform the participants in advance that they were going to be interviewed about their planning strategies and consequently this did not lead to problems of reactive effects. This suggests that learners may be given a broad indication of the study's aims but not specific. The participants in the present study were therefore told that they were going to be interviewed after performing a task but they were not informed about planning strategies or the targeted RC types thus minimising reactive affects. Furthermore, as the procedure was repeated in weeks three and four, the students may well have expected post-task interviews to occur, but as there

was no explicit indication about the topic of the interview, reactive affects were minimized and the interview data was comparable throughout the treatment.

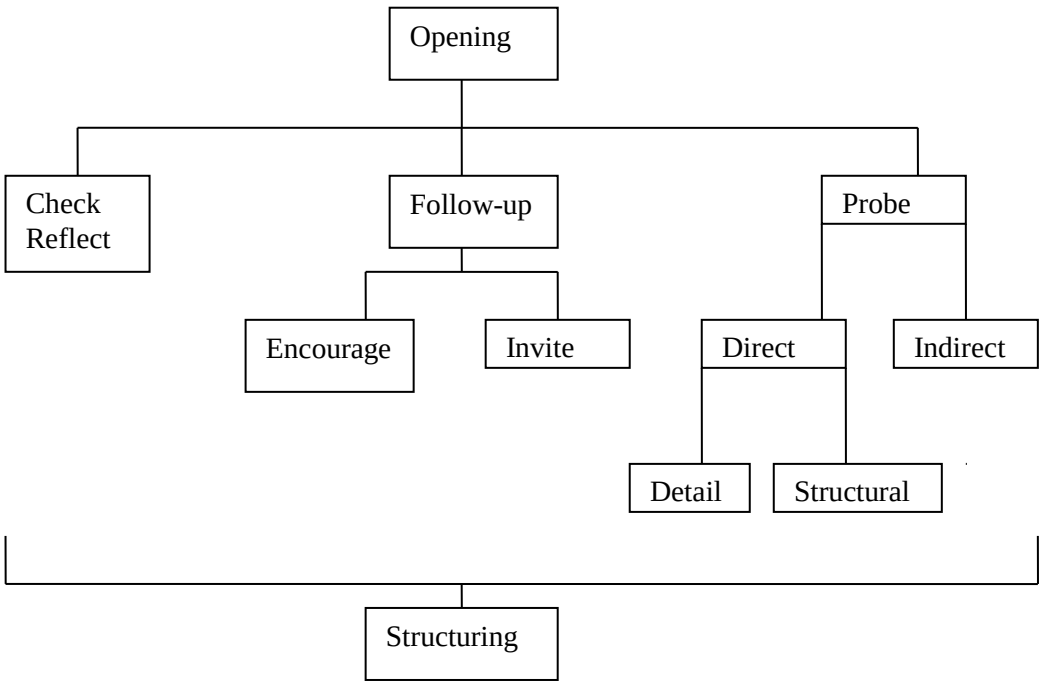
5. The argument that learners cannot successfully recount previous thought processes because such processes are unconscious can be addressed in the following way. Dornyei (2007) suggests using a stimulus to help learners access their thoughts. Gass & Mackey (2000) argue that a visual reminder of the event will help participants retrieve their thoughts and comment on them. In the present study, the GP and GUP groups were instructed to make notes whilst they planned for a task. One option for a visual reminder would be to video record the learners as they planned for a task and then use the recording as a stimulus for students to watch and comment on their planning strategies. However, as this study was focusing on planning rather than task performance, it was considered that the students' planning notes would be a greater stimulus to help recall learners' planning strategies rather than watching a video of themselves planning. In addition, previous studies such as Ortega (1999) used students' planning notes as a stimulus during interviews to help retrieve thought processes about planning. Consequently, the present study carried out stimulated recall interviews using learners' planning notes as a visual reminder instead of watching a video. Each learner was provided with their planning note-sheet at the start of each interview which they could refer to when attempting to describe their planning strategies.

To conclude, stimulated recall does not come without flaws. However, this technique is particularly useful for the purpose of this study as it can enable us to gain valuable insights into how learners prepare themselves to speak in the L2, which we saw in the results of the pilot study in 4.5.6 and 4.5.7. For example, how the B2 guided planners continually attended to form as they planned for more complex tasks over time. A major factor in the success of the pilot study in being able to elicit learners' strategies was the steps taken to minimize the disadvantages of stimulated recall outlined above. For example, the interviews were always carried out immediately after learners had planned and performed a task, and they were always provided with their planning notes which helped them recall their planning strategies. As a result, the present study employed retrospective interviews in order to investigate what learners did when they planned whilst following the above steps to minimize the technique's potential flaws.

5.6.1 Interview question types

To answer this study's second research question, a series of interview questions were used from our pilot study to enable participants to engage in a process of metacognitive retrospection. In order to obtain detailed information from the interviewee, open questions are recommended as they prompt the participants to explain their responses (Robson, 2002). In the case of this study, one could ask: *'how did you plan for this task?'* On the other hand, Richards (2003) also points to the value of closed questions in order to confirm a certain point, for example, yes/no responses. According to Richards (2003), there are no set rules about the types of questions that should be asked in an interview as each one is unique depending on the purpose and how the participant responds to the questions asked. Richards (2003) does recommend using different question types at various stages of an interview in order to obtain the required information (see figure 11).

Figure 11. *Interview question types (Richards, 2003, p.57)*



Robson (2002) refers to this type of interview as a "semi-structured interview" (p. 270) which is commonly used within qualitative research as the order of questions "can be modified based on the interviewer's perception of what seems most appropriate.

Question wording can be changed and explanations given; particular questions which seem inappropriate with a particular interview can be omitted".

Richards (2003) recommends starting an interview with an opening question which attempts to elicit a long response. The question should not be specific but should be general enough that it "provides a natural springboard for further questions" (p. 56). Robson (2002) suggests a similar approach which he refers to as "warm-up' easy, non-threatening questions at the beginning to settle you both down" (p. 277). In Ortega (1999) interviews began after task performance with questions such as "how was it?" (p. 148) which Ortega (1999) referred to as 'ice-breaker' questions. As a result, the researcher started each interview using this type of question, before moving onto more specific questions such as *'how did you plan for this task?'* Richards (2003) notes that if the interviewee's response is unclear in any way "it's always worth checking this or reflecting a statement back to the speaker. This may be achieved using a "check reflect" (p. 56) question which may also prompt the speaker to develop a point further. In the case of this study it could be something like *'Did you say you focused on grammar?'*

Follow-up questions can be used "when the speaker has raised something or perhaps given a subtle indication that there is more to be discovered on this topic" (Richards, 2003, p. 56). In order for the interviewer to obtain further information on the matter, simple encouragement may do, or the interviewer may invite the participant to expand on a certain point. For example, a hypothetical follow-up question to a learner's planning strategy could be *'You said you focus on vocabulary, do you always focus on vocabulary?'* According to Richards (2003) "points will emerge during the interview that demand more careful excavation and here the interviewer will need to probe specific elements in order to build up a satisfactory picture (p. 56)." In order to obtain more detail to an interviewee's response, 'wh' questions are the most direct method. For example, as we are interested in how learners' planning strategies may change over time, we could ask *'Did you plan differently compared to last week?* If the answer is yes, we could reply with *'In what way?'* Ortega (1999) used a series of 'wh' questions to probe for further information, for example, "what do you mean by that?" (p. 148).

"Finally, in a formal interview, it may be necessary to mark a shift of topic by using structural moves such as "Can we move onto…" (Richards, 2003, p. 57). These 'structuring' questions are useful in maintaining the pace of an interview, enabling the interviewer to remain in control of the interview by discussing what they feel is most important.

So far we have discussed the different types of questions that can be used in an interview. Indeed, Ortega's (1999) study consisted of semi-structured interviews using an array of different question types which were adapted and used in the pilot study. The next step was to create an interview design that allowed us to structure questions in the correct order so that we could successfully obtain responses from the participants.

5.6.2 Interview design

Richards (2003) provides a series of steps to consider for preparing an effective interview:

1. "Decide on what the interview is setting out to achieve" (p. 69). Having a clear understanding of the purpose of the interview will help to formulate appropriate interview questions. In the case of this study the purpose involves learners' planning strategies for oral tasks that increase in complexity and how these strategies may change over time.
2. "Identify the big questions" (Richards, 2003, p. 69). This can be achieved by formulating interview questions based off the research questions or the main topics of investigation. In the case of research question two: what strategies do Japanese second year university learners of English use when planning for oral narratives that increase in complexity over time? The main topics to consider are:

(a) The planning strategies of Japanese learners for oral tasks.
(b) How these strategies may change as tasks increase in complexity over time.

Thus, after a general warm-up question, the main interview questions were:

1. *'Did you plan for this task?'*
2. *'How did you plan for it?'*
3. *'What was your focus when you were planning?*
4. *'Did you think about grammar, vocabulary, how to organise the story, or something else?'*

Then as each week progressed, the following main questions were also used to find out whether the learners' planning strategies changed:

5. 'Did you plan differently compared to last week?'

6. 'In what way?'

Richards (2003) points to the advantage of having subsequent interviews as "there will always be issues arising from the first interview, some of them only revealed by subsequent analysis, and a second meeting will provide an invaluable opportunity to develop lines of investigation" (p. 69). This approach was used in the present study as the researcher could reflect on learners' responses during one week and then probe certain topics or issues the following week in order to obtain further information.

3. "Decide on lines of enquiry deriving from these" (Richards, 2003, p. 70). This involves formulating subsidiary topics that come under the main topics and involve sub-sets of questions to elicit further responses. For example, if the guided planners mentioned that they focused on grammar, a series of reactive questions taken from Ortega (1999, p. 148) were used to elicit further information such as *'can you give me an example?', 'what do you mean?'*

4. "Analyse, apply, review, revise" (Richards, 2003, p. 70). In order to elicit the desired responses from participants, Richards (2003) stresses the need to analyse the interview guide from the participant's perspective so as to identify possible difficulties. The interview should then be piloted, reviewed and amended with any necessary adjustments. As we saw in the last chapter, the interview questions adapted from Ortega (1999) for this study were also used in the pilot study. As mentioned in 4.6, the questions successfully elicited the planning strategies of the B1 and B2 learners over time, and as a result, only minor alterations to the questions were considered necessary for the main study. For example, shortening the questions to make them more simple, for example *'could you explain how you planned for this task?'* to *'how did you plan?'* See appendix O for the list of questions used.

The final factor concerning the interview questions was the language used to convey them. Dornyei (2007) recommends that interviews should be carried out in the participant's L1, as in Ortega's (1999) study. In the present study, the researcher was unable to conduct interviews in Japanese or use a Japanese person to carry out the interviews and transcribe the data. As a result, the researcher had to conduct the interviews in English. In order to test whether it would be possible to carry out interviews regarding learners' planning strategies in the L2, the questions were trailed during the pilot study. The results of the pilot study were pleasing as the B1 and B2

learners were able to provide detailed responses regarding their planning strategies in English, as can be seen in their responses in 4.5.6 and 4.5.7, and in their interview transcriptions in appendix I. Given the satisfactory responses from the intermediate participants in pilot study, the researcher was confident that the intermediate participants in the main study would have sufficient L2 ability to be able to respond to the interview questions. Although it was expected that the main study participants would not provide as much detail regarding their planning strategies compared to the pilot study learners because of the contrasting university backgrounds of both sets of students. As discussed in 4.3.1, the participants in pilot study were studying at APU which is one of the few bilingual Universities in Japan and all the students study English proficiency TOEFL exams. Thus, students at APU receive much greater exposure using English in different situations compared to the students in the main study who were studying Economics and Business at Shimonoseki City University and did not have TOEFL scores. Consequently, although both sets of learners were recruited from intermediate level English classes, the interview data obtained from the main study was expected to be limited due to the fact that the participants had much less exposure using English on university campus compared to the participants in the pilot study. As a result, performing an interview in English about planning strategies would be a more difficult task for the main study participants and this would compromise the quality of the interview data compared to the pilot study. To help overcome this issue, the participants were provided with their planning note-sheets during the interviews which were expected to help them remember their planning strategies and respond in the L2, as the students' notes were written in English. The limitations of the interview data for the main study will be further discussed in 9.5.

5.6.3 Interview procedure

As outlined in 5.5, the interviews took place during weeks two, three and four of the study. During each week, students would meet the researcher for approximately one hour in a private classroom outside of regular class time. Students were scheduled to meet the researcher in pairs or in groups of three or four. Each student would first take turns planning and performing one narrative task one-on-one with the researcher who would act as the listener. The other students would wait outside the classroom. After

which, each student was interviewed one-one-one with the researcher regarding their planning strategies whilst the other students waited outside the classroom.

No photographs were taken of the interview setting, however the classroom itself was a standard small to medium sized room holding approximately thirty students and it contained a whiteboard and a TV. The researcher would sit with each student at a desk, then provide the student with his/her planning note-sheet and then begin the interview. The researcher asked the questions described in the last sub-section in English and recorded the interview. During week two, the average interview lasted 3.68 minutes. At week three, the average interview lasted 2.37 minutes and at week four the average interview lasted 2.40 minutes. A limitation of this form of data collection concerns the length of the interviews as only a limited amount of data was obtained per student concerning their planning strategies. Ortega (2005) reported that her interviews lasted approximately 10 minutes per participant for each task which are longer than the interviews in the present study. Although additional questions could have been used to probe students' responses for a longer period of time, the study was constrained by a time limit. Each student was paid by the hour to plan and perform a task, complete an interview and then plan and perform one more task. In order to meet all the students within the designated time-frames, the researcher had to co-ordinate time-slots within each hour for each student to complete the tasks and interviews. This resulted in a time limit for the interviews which could not be extended beyond six minutes per student. This limitation would have compromised the quality of the data due to the lack of information obtained regarding each learner's planning strategies. The impact this had on the study's results will be discussed in 9.5.

Having discussed and justified the use of post-tasks interviews to help us answer research question two, the next sub-section provides a similar account for the use of a pre-treament questionnaire.

5.6.4 Pre-treatment questionnaire

A pre-treatment questionnaire was included in the main study to investigate learners' perceptions towards L2 speaking to see whether individual differences existed in terms preferences towards communication or accuracy. As research question two was interested in comparing the planning strategies of two groups under different planning

conditions: one that was form-focused (GP) and the other that was less form-focused (GUP), it was considered important that both groups were considered equal in their preferences towards oral communication and accuracy prior to engaging in their respective planning treatments. However, given the time cost involved in administrating and analysing our interview data, a research instrument was needed that would be relatively quick to design, administer and analyze, hence the use of a questionnaire. The format of the questionnaire was adapted from Timmis (2003) in which participants were asked to respond to hypothetical statements about speaking in the L2. The questionnaire consisted of just two items, each one contained a statement that was adapted from the post-task interviews in Ortega (2005):

1. "*I often think about the errors I make when speaking English and sometimes I don't want to speak if I make a mistake. I'm never happy with my grammar."*

2. "*I think communication is more important than grammar. When learning to speak English, it's natural to make mistakes but through practice we can improve."*

As we saw when we reviewed Ortega's (2005) study in 3.3.7, statement one indicates a preference towards accuracy whilst statement two shows a preference towards communication. The participants were required to rate each item on a scale of 1 to 7 (1 – strongly disagree, 2 – disagree, 3 – slightly disagree, 4 – unsure, 5 – slightly agree, 6 – agree, 7 – strongly agree) (see appendix P).

Oppenheim (2000) notes that questionnaires are easy to administer and answer, and they involve low cost in terms of data collection and processing. Thus, a questionnaire appeared to suit our needs as a simple instrument to use for this particular aspect of the study given the time expense for data collection and analysis involved with our interviews. Indeed Oppenheim (2000) mentions that questionnaires and interviews compliment each other in terms of advantages and disadvantages. For example, as we saw in 5.6, interviews allow us to clarify any misunderstood points, however, this is a disadvantage with questionnaires as there is no opportunity to probe, or correct misunderstandings. Questionnaires may also be unsuitable for learners of low L2 proficiency who may not understand the items. Finally, Cohen et al. (2005) point out that "most of us would not like to be called extremists" (p. 254) and as a result,

participants may be reluctant to use the extreme values at either end of a scale. As a result, rating scales may not provide enough range to give accurate responses.

Despite the disadvantages to questionnaires, the researcher believed they could be minimized by the design of the overall study. For example, any misunderstandings or the need to probe learners' responses regarding their L2 orientation could be addressed during the subsequent post-task interviews which show the benefits of a mixed methods approach that we discussed in 4.2, in which one form of data collection can support the findings of another. In terms of language difficulty within the questionnaire, the researcher considered the language to be simple enough for intermediate learners to comprehend given his several years teaching experience with learners of similar proficiency. In addition, the questionnaire was pre-piloted on an intermediate student who did not participate in the study but was in the same class as students who did. The student was able to comprehend and complete the questionnaire, and as a result, no amendments were made to the items. Finally, in order to provide more range for students' responses, the questionnaire was designed to include seven response items instead of the five listed in Timmis (2003). Thus if students were reluctant to use the extreme values, they still had five items to choose from.

To conclude, although there are disadvantages to using questionnaires, the present study took steps to address those issues and as a result, it was considered that the inclusion of the pre-study questionnaire would add more pedagogic value to the findings of the main study that was lacking in the pilot study. For example, the questionnaire results would enable us to see whether differences existed between the two groups in terms of their preferences towards speaking in the L2 prior to engaging in their respective planning conditions. We could then see how much of an effect each planning condition had on each group's planning strategies in relation to their personal preferences towards L2 speaking.

Table 24 displays the results of the GP group and table 25 displays the results of the GUP group. They provide each participant's identification number, the response to statement one and two, followed by an interpretation of their L2 oral orientation.

Table 24. *GP group's oral orientation towards communication and accuracy*

GP Group	Statement 1: Accuracy Orientated	Statement 2: Communication Orientated	L2 oral orientation
101	6	5	Accuracy and Communication
102	2	7	Communication
103	3	5	Communication
104	5	5	Accuracy and Communication
105	6	7	Accuracy and Communication
106	4	6	Communication
107	3	5	Communication
108	4	6	Communication
109	6	6	Accuracy and Communication
110	6	5	Accuracy and Communication
111	2	7	Communication
112	5	5	Accuracy and Communication
113	4	5	Communication

The results showed that individual differences existed within the GP group regarding their oral orientation towards communication and accuracy. For example, 7 out of the 13 GP students (54%) rated a preference towards communication compared to accuracy when speaking in the L2. None of the participants indicated they were more orientated towards accuracy over communication, however, 6 participants (46%) indicated a preference towards communication *and* accuracy when speaking in the L2. We can therefore assume that approximately half of the GP group preferred to focus on meaning and fluency when speaking in the L2, whilst half the participants showed an interest attending to form *and* meaning whilst speaking in the L2 so as to speak accurately *and* fluently. The results of the GUP group's pre-study questionnaire are displayed below.

Table 25. *GUP group's oral orientation towards communication and accuracy*

GUP Group	Statement 1: Accuracy Orientated	Statement 2: Communication Orientated	L2 oral orientation
214	4	6	Communication
215	2	6	Communication
216	3	6	Communication
217	5	6	Accuracy and Communication
218	5	6	Accuracy and Communication
219	6	5	Accuracy and Communication
220	5	6	Accuracy and Communication
221	3	6	Communication
222	5	6	Accuracy and Communication
223	3	6	Communication
224	6	6	Accuracy and Communication
225	3	5	Communication
226	5	7	Accuracy and Communication

These results showed that individual differences also existed within the GUP group regarding their oral orientation towards communication and accuracy. Six out of the 13 GUP learners (46%) rated a preference towards communication compared to accuracy when speaking in the L2. None of the participants indicated they were more orientated towards accuracy over communication, however, seven participants (54%) indicated a preference towards communication *and* accuracy. Approximately half of the GUP learners therefore preferred to focus on meaning and fluency when speaking in the L2, and half the participants showed an interest attending to form and meaning so as to speak accurately and fluently.

From these results we can see that the GP and GUP groups appear to be similar in their L2 orientation towards speaking. Approximately half the learners in each group preferred to focus on communication and meaning, whilst half the learners showed an interest attending to form and meaning whilst speaking in the L2 so as to speak accurately *and* fluently. An independent samples *t*-test was carried out to determine whether there were any significant differences between the GP and GUP groups in

terms of their oral orientation towards L2 speaking. The results showed that the GP group were slightly more orientated towards accuracy (M = 4.31, SD = 1.49) compared to the GUP group (M = 4.23, SD = 1.30) but this difference was not significant at the .05 level, $t(24) = 0.140$, $p > 0.05$. In addition, the GUP group were only slightly more orientated towards communication (M = 5.92, SD = .49) compared to the GP group (M = 5.69, SD = .85) but again, this difference was not significant $t(19.2) = -.843$, $p > 0.05$. Consequently, both groups were considered equal in terms of their L2 oral orientation towards communication and accuracy prior to engaging in their respective task planning treatments.

After describing and justifying all the materials necessary to carry out the main study, we now turn to the analysis methods of the data.

5.7 Data Analysis

This section will describe the data analysis methods concerning research question one and two. We begin by describing and justifying the measures used to answer research question one: namely the quantitative measures relating to accuracy, syntactic complexity, fluency and the grammatical judgement test that were all used to assess L2 development. We then move onto describe and justify the analysis measures used for research question two: specifically the qualitative measures relating to the post-task interviews. Finally, the computer software (CLAN) that was used to code and analyze the measures relating to accuracy, syntactic complexity, fluency as well as the post-task interviews will be explained and justified in sub-section 5.7.6.

5.7.1 Quantitative data analysis: L2 oral development

To examine the effects of guided planning and task complexity (GP) compared to guided and unguided planning and task complexity (GUP) on L2 oral development, the analysis of hypotheses one to four was carried out in two main steps. First, the pre-post-test gains within each group were analyzed in the following way:

- Differences between the pre- and immediate post-test scores
- Differences between the pre- and delayed post-test scores
- Differences between the immediate and delayed post-test scores

As pointed out in the pilot study (4.3), Schmitt (2010) notes that data analysis regarding pre- to immediate post-tests helps "to determine the effect of the treatment" (p. 155) whilst pre- to delayed post-tests, as well as immediate to delayed post-tests "can demonstrate if long-term retention (i.e learning) has occurred" (p. 156).

Paired samples *t*-tests were carried out to calculate differences in mean sizes from the pre- and post-tests within the GP and GUP groups. Paired samples *t*-tests are commonly used to test for significant differences within a group's performance at different points in time (Field, 2009). The level of significance was targeted at $p < 0.05$ which is the standard level for these types of analyses (Dornyei, 2007). Thus, any pre-post-test gains would be considered statistically significant at $p < 0.05$.

Following Mochizuki & Ortega (2008), another important form of analysis involved the effect sizes of the planning groups. Effect sizes "provide information about the magnitude of an observed phenomenon" (Dornyei, 2007, p. 212). They are calculated using a standard measure, the advantage being that they allow us "to compare the results reported in different studies, because the effect size indices are transformations onto a common scale". Mochizuki & Ortega (2008) used Cohen's (1988) *d* value to calculate effect sizes of their planning conditions. Cohen's (1988) *d* value is based on a mathematical formula which compares the means of two samples with their standard deviations. According to Cohen (1988), if a *d* value is approximately 0.20, the effect between the two samples is small, whilst 0.50 shows a medium effect, and approximately 0.80 or above is a large effect. Cohen's (1988) *d* value also has the advantage of being a preferred measure when data sets contain non-normal distribution (Field, 2009). Thus, in addition to the paired samples *t*-tests, *d* values were calculated to obtain the effect-sizes of the pre- post-test mean scores of the GP group and the GUP group.

Step two of our analysis determined which type of task treatment provided optimal results for L2 development. To do this, comparisons were made *between* the GP group and the GUP group using independent samples *t*-tests. An additional benefit of this *t*-test for the purpose of this study was that it accounts for non-normality when comparing two sets of data as it provides corresponding adjustment values (Field, 2009). Cohen's (1988) *d* value was also used for effect sizes between the two groups.

Appendix Q provides the transcriptions of two samples of the pre- and post-test narration performances.

5.7.2 Accuracy of OS and OPREP RC types and 3rd person singular and plural

As mentioned in 2.4.3, our definition of accuracy relates to "the extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm (i.e. usually the native speaker). Thus, deviations from targetlike performance would be considered errors" (Housen et al., 2012, p. 4). In addition, accuracy also relates to "*appropriateness* and *acceptability*" (p. 4). For the purpose of this study, deviations from targetlike performance relate to grammatical errors and communicatively inadequate use of the targeted forms: OS and OPREP RC types, 3rd person singular and plural. In other words, if a learner produced an OS RC type that was grammatically correct but it did not reflect the context of the storyline it would be considered inaccurate. It was therefore necessary to use a measure that could gauge grammatical errors relating to OS and OPREP RC types and the accompanying use of 3rd person singular or plural as well as communicatively inadequate use of the forms in relation to targetlike performance. For the purpose of this study, Mochizuki & Ortega's (2008) rating scale that was designed to measure oral accuracy of simple RC types was adapted and used. As described in 3.3.6, Mochizuki & Ortega (2008) designed a 6 point rating scale in which each point represented a category of grammatical accuracy in a spoken context. For example, five points were awarded for targetlike relativization in which the relative pronoun was used correctly. Points were then reduced for grammatical errors relating to the RC type with 0 points awarded as the lowest score which indicated avoidance of the form. Table 26 (section a) provides an illustration of the rating scale used in Mochizuki & Ortega (2008) including definitions and examples regarding the six points relating to RC production.

Table 26. *Relative clause and relevant morphology rating scale (adapted from Mochizuki & Ortega, 2008, p. 22)*

Section B: Grammaticization: Morphological Adequacy Scale

Descriptor	Definition	Example	Points
Target-like use of 3rd person singular & targetlike RC	Two instances of targetlike use of of 3rd person singular that accompany targetlike relativization involving 3rd person singular or plural	*'He thinks that he likes the dog which has long ears'* *'He thinks he likes the dogs which have long ears'*	9
Target-like use of 3rd person singular & targetlike RC	One instance of targetlike use of 3rd person singular that accompany targetlike relativization involving 3rd person singular or plural	*'He wants the dog which has long ears'* *'He wants the dogs which have long ears'*	8
Target-like suppliance of RC only	Use of 3rd person singular that contain errors that compliment a targetlike RC that contains no errors	*'She thinks he like the dog which has long ears'* '*He thinks like the dog which the woman is looking at'*	7
Target-like suppliance	A relative clause that exhibits targetlike relativization; contains no errors relating to verb tense but may contain other errors such as articles	'*He want the dog which has long ear'* '*He want the dogs which have long ear'*	6

Section A: Syntacticization: Relative Clause scoring scheme

Descriptor	Definition	Example	Points
Target-like suppliance	A relative clause that exhibits targetlike relativization; it may contain one or more errors that are irrelevant to the target structure, such as verb tense or	*'I want the dog which have long ear'*	5

	the use of articles		
Developmental suppliance	A relative clause that contains any of four error types (i.e. pronoun rentention nonadjecency, incorrect relative marker and inappropriate relative pronoun omission) described in the previous studies on relative clauses (e.g. Izumi, 2003)	*'I want the dog which many people are watching dog.'* *2. 'The dog is friendly which has long hair.'* *3. 'Ken likes the dog who has long ears.'* *4. 'I like the dog has long ears.'*	4
Attempt with processing overload	Relative clause attempted but containing a breakdown such as omission of head noun or verb in the relative clause.	*'She wants which has long ears.'* *'She wants the dog which long ears.'*	3
Least successful attempt	Relative clause where both developmental and processing load errors combine to cloud the success of the product and hinder intelligibility	*'Kanako wants to buy which has long hair and long ear dog.'*	2
Simplification	An utterance in which the participant tried to convey meaning without attempting relativization, alternative structures; these include either the structure derived from a direct translation form Japanese or alternative structures in English	*1.'long the dog that has long ear.'* *2. 'the dog with long hair.*'	1
Avoidance of Content	Formulation of the content involved in one of the seven contexts for obligatory suppliance was not attempted		0

According to Mochizuki & Ortega, the six point scoring system was designed because the participants in their study "not only were of low proficiency but also had little familiarity with speaking" (p. 21). Consequently, the rating scale allowed for a more sensitive measurement of RC accuracy compared to previous measures that involved binary accuracy scoring such as 'error-free clauses'. The latter measure was considered unsuitable for the beginner level participants of the study who may not have had the ability to produce error-free instances of the form.

Despite the intermediate proficiency of the present study's participants, the rating scale was still used as a measurement of accuracy for three reasons. First, as we saw in 4.5.1, the accuracy measures used in the pilot study were 'error-free relative clauses per AS-unit' and 'error-free relative clauses per relative clause'. However, given the all-or-nothing binary feature of these measures, not many cases of error-free RCs were produced by the B2 learners. The advantage of the rating scale used in Mochizuki & Ortega (2008) was that it allowed for more sensitive improvements in learners' grammatical development of RC use rather than accepting only a holistically accurate RC. Second, Mochizuki & Ortega's (2008) narrative consisted of seven obligatory contexts of RCs, thus their rating scale "maintains the number of contexts to be assessed constant across learners (k = 7), thus making the scores into a true interval scale that can be directly submitted to referential analysis" (p. 21). In other words, the maximum score achievable would be 35 points (5 points multiplied by the 7 contexts) and this would remain constant for all participants. As a result, scores would not need transforming for conformity and could be directly inputted for inferential statistical analysis. Finally, the rating scale would be particularly useful when comparing pre- and post-test scores as it would allow us to see any potential gains made as a result of the treatment. For the purpose of this study, the rating scale used in Mochizuki & Ortega (2008) was adapted to incorporate targetlike relativization of the OS and OPREP RC types, 3rd person plural and the accompanying use of 3rd person singular. This involved adding an additional four categories at the top of the rating scale that were categorized under section B, see table 26. Points six and seven relate to targetlike relativization including correct use of RCs that were located next to singular or plural head nouns thus reflecting correct use of 3rd person singular or plural respectively. Points eight and nine however, concern target-like use of 3rd person singular within the independent clause as well as target-like use of the adjoining RC type. As a result, the rating scale designed for this study incorporates two sections of grammatical accuracy. Section A concerns syntacticization i.e. syntactic

adequacy of RCs, whilst section B relates to grammaticization i.e. morphological adequacy of RCs and the complimentary use of 3rd person singular. Ellis & Barkhuizen (2005) note that targetlike oral morphology is appropriate for syntactic accuracy when using focused tasks designed to elicit specific grammatical forms, as in the case of this study. Consequently, target-like morphology for the present study is a specific measure of accuracy. The maximum score achievable for the 7 obligatory RC contexts in each pre- and post-test using our adapted rating scale was 60 points (see figure 12 for an illustration of possible phrases to be used in the pre-test narrative and the points awarded).

Figure 12. *Pre-test narrative: possible examples for maximum points for each RC context*

Context 1 (OS): The mother thinks that she likes the dog which has long ears	9
Context 2 (OS): She wants the dog which is next to the girl	8
Context 3 (OS): Kevin thinks he likes the dogs which have long hair	9
Content 4 (OPREP): He wants the dog which the family is looking at	8
Context 5 (OS): Kate thinks she likes the dog which has long hair	9
Context 6 (OS): She also thinks she likes the dog which has long ears	9
Context 7 (OPREP): She wants the dog which the girl is smiling at	8
Maximum total:	60

The uneven scoring of each RC context was due to the extent to which 3rd person singular was required to accompany each RC type. For example, contexts 1, 3, 5 and 6 could be awarded 9 points as they facilitated two instances of 3rd person singular that accompanied targetlike use of the RC type as they involved a character *thinking* about something represented by thought bubbles, for example, *'he thinks that he likes a dog which has long hair'*. The remaining contexts involved a character *choosing* something and therefore required only one use of 3rd person singular that accompanied a targetlike RC, for example, *'he wants the dog which has long hair'*, and as a result would be awarded 8 points according to the rating scale.

In terms of the remaining part of our accuracy definition: communicatively adequate use of the targeted forms, learners' accuracy was based on the context in which they used the forms. In other words, the learners were required to produce seven

obligatory contexts of RCs that reflected the storyline, thus if an RC was produced that was grammatically correct but communicatively inadequate i.e. it did not relate to the storyline, it would not receive a score. For example, describing a picture that contains a dog with long ears but commenting *'the cat which has short ears'.* However, if the learner immediately self-corrected their use of the RC type then he/she would be graded on the self-correction. Thus, in line with the analysis of the pilot study in 4.4.1, repeated RCs, self-corrections and false starts were excluded from the analysis. For example, as in the underlined structures '*he likes the dog <u>which has black hairs</u>, ah, which has black hair*.' In addition, if a learner described an obligatory context using a different RC type, for example, using the OS in context 7 instead of the OPREP, they would not be penalized providing its use was appropriate, in other words it reflected the storyline.

Finally, in line with Mochizuki & Ortega (2008), this study prevented 'over-use' of the targeted forms by setting a maximum score achievable for all participants. In Mochizuki & Ortega's (2008) study "maximum total score in the use of relativization in the narrative task was 35 points (a maximum of 5 points by 7 contexts for obligatory suppliance" (p. 21). In the case of the present study, it was 60 points, as outlined above. Thus, if a learner produced twelve RCs during a narration, they would only be graded on the seven contexts as shown in figure 12. As in Mochizuki & Ortega (2008) this score would remain constant for all participants when comparing the pre- and post-test narratives scores. For instructions of how accuracy was coded and analysed using the computer software CLAN see appendix R.

The advantages of rating scales are that they involve the use of closed questions in which a range of responses are provided for an individual to respond to (Cohen et al. 2005). In other words, an individual is simply required to choose the most appropriate item. As a result, rating scales with closed questions "are quick to complete and straightforward to code (e.g. for computer analysis) and do not discriminate unduly on the basis of how articulate the respondents are" (p. 248). Cohen et al. (2005) describe that rating scales are a popular form of measurement in research as "they combine the opportunity for a flexible response with the ability to determine frequencies, correlations and other forms of quantitative analysis" (p. 253). Furthermore, the reliability of rating scales can be tested by comparing the scores of multiple assessors. For example, if different assessors can grade a student's production of RCs with similar consistency then the rating scale can be considered a reliable construct for testing learners' oral performance of the form. This procedure, otherwise known as interator

reliability was carried out on the present study's rating scale by following a similar process used in Mochizuki & Ortega (2008). Two independent rators who were both experienced EFL teachers in Japan were asked to rate a random 10% sample of the data taken from the pre- and post-tests. This consisted of scoring eight student narratives (two from the pre-test, three from the immediate post-test and three from the delayed post-test). Overall, both rators showed consistency in their responses to the accuracy of the targeted forms as on average, the scores were the same or differed by +/- 1 point per context (see appendix S). The results of independent samples *t*-tests also showed there were no significant differences between the means of both rator's for each of the eight narratives as shown in table 27. As a result, the researcher was satisfied with the level of interator reliability using the present study's accuracy rating scale.

Table 27: Interator reliability significance values between two independent rators

Narratives	Rator 1 Mean	Rator 2 Mean	Significance value
Narrative 1	1	1	N/A as both rators averaged ($M = 1$)
Narrative 2	1.57	1.43	$t(12) = .200$, $p > 0.05$
Narrative 3	6.43	6.57	$t(12) = -.098$, $p > 0.05$
Narrative 4	4.71	4.86	$t(12) = -.090$, $p > 0.05$
Narrative 5	6.57	6.86	$t(12) = -.195$, $p > 0.05$
Narrative 6	5.14	5.14	$t(12) = .000$, $p > 0.05$
Narrative 7	5.86	7.00	$t(12) = -.691$, $p > 0.05$
Narrative 8	6.00	5.14	$t(12) = .679$, $p > 0.05$

In terms of disadvantages of rating scales, Cohen et al. (2005) point out that "there is no assumption of equal intervals between categories, hence a rating of 4 indicates neither that it is twice as powerful as 2 nor that it is twice as strongly felt" (p. 254). Thus, if student 'a' produced one RC type that was graded '4', and student 'b' produced one RC type that was graded a '2', we could not infer that student 'a's production of RCs was twice as accurate as student 'b'. Cohen et al. (2005) also mention "there is no check on whether the respondents are telling the truth" (p. 254). Finally, as with questionnaires, rators may be reluctant to grade the extreme values at either end of the scale. Consequently, a six point rating scale, as in Mochizuki & Ortega

(2008), would only offer a choice of four responses which does not provide much range to give accurate responses.

In the case of the present study however, all these points were addressed. Firstly, this study acknowledges that the rating scale does not have equal intervals between each category, only that each category increases in grammatical accuracy of the form. Thus, a grade of 4 will not be acknowledged as twice as accurate as a grade of 2, but it will be acknowledged as being more grammatically accurate. In terms of respondents telling the truth, this issue was addressed through our use of interator reliability in which both rators provided similar scores thus indicating a true reflection of the students' performance. Finally, the extreme values were addressed in a similar way to the questionnaire in that additional items were added to the rating scale so that it consisted of nine responses. This allows for a seven point variation after omitting the two extreme values which provides sufficient range for rators to accurately assess learners' performance. To conclude, although the accuracy rating scale measure may be limited, for the purpose of this study it does allow us to measure grammatical accuracy and communicatively adequate use of the OS, OPREP RC types, and 3rd person singular and plural in line with our definition. We now discuss the complexity measure of the present study.

5.7.3 Syntactic complexity

In 2.4.5 we defined syntactic complexity as:

> a language feature or system of features is seen as complex if it is somehow costly or taxing for language users and learners, particularly in terms of the mental effort or resources that they have to invest in processing or internalizing the feature(s). (Bulte & Housen, 2012, p. 23)

For the purpose of this study, syntactic complexity refers to English RCs which are a grammatical feature known for its cognitive difficulty in L2 oral production and development with Japanese learners. It was therefore important to use a measure that reflected our definition. The measures used in the pilot study were 'relative clauses per AS-unit' and 'dependent clauses per AS-unit'. As the present study involved a larger sample size that would require more time for data analysis, as well as the extra time needed to analyse the data from the grammatical judgement test, it was decided that

syntactic complexity could only comprise of one measure. Consequently 'relative clauses per AS-unit' was chosen as it was the only measure that targeted the form of the present study: RCs, which as we know from 3.5.4, is a grammatical feature known for its cognitive difficulty in oral production with Japanese learners of intermediate proficiency. As the purpose of this study was to develop Japanese learners' use of OS and OPREP RC types 'relative clauses per AS-unit' was therefore considered a suitable measure that would reflect our definition of syntactic complexity. Furthermore, in the pilot study (4.5.3) the measure reported positive gains as a result of guided planning and task complexity which further justified its inclusion in the main study.

As 'relative clauses per AS-unit' was used in the pilot study, all examples of the measure that was explained in 4.4.2 (and in 4.4.1 in terms of AS-units) apply to the main study analysis. For example, following Mochizuki & Ortega (2008), non-target like use of RCs were accepted providing the learner used a relative pronoun. Repeated RCs were excluded from the analysis in order to prevent over-use of the form, as in the underlined structure 'he likes the dog *which has black hair*, *which has black hair*'. AS-units that contained grammatical errors were accepted provided they had pragmatic meaning however repeated clauses were excluded from the analysis. Following Mochizuki & Ortega (2008), and Gilabert (2007a), syntactic complexity was measured by totalling the number of RCs within each narrative and then dividing them by the number of AS-units. For instructions of how syntactic complexity was coded and analysed using the computer software CLAN, see appendix T.

As mentioned in 2.4.6, criticisms have been levelled at syntactic complexity measures based on the assumption that the construct is supposed to consist of measures that involve aspects of subordination that are considered difficult for learners to use. Bulte & Housen (2012) also comment on the adequacy of syntactic complexity measures as reliable tools for indicating *development.* "Any measure that serves as an index of development would probably have to cover the full trajectory of language acquisition, from the lowest level or stage to the highest" (2012, p. 37). Thus, an investigation into the development of a linguistic feature would need to use a syntactic measure that could capture all types of the feature both simple and complex. Finally, "limitations of subordination measures as indices of syntactic or of more general grammatical development mainly stem from their specific and fairly narrow linguistic scope" (Bulte & Housen, 2012, p. 37). This is because subordination measures only represent one form of syntactic complexity: the sentential level and not the clausal or

phrasal level of syntactic complexity. In other words, subordination measures focus on one form of syntactic complexity: the embedding of subordinate clauses. Other forms of syntactic complexity such as coordination, pre- and post modification of noun phrases, and nominalisation are not represented by subordination measures. Thus, subordination measures are limited in terms of representing a full range linguistic complexity.

Despite the disadvantages of subordinate measures as an indicator of syntactic complexity, for the purpose of this study, they are still warranted for the following reasons. In terms of subordination reflecting difficulty, we know that RCs are considered to be a difficult linguistic feature for intermediate Japanese learners to acquire given the differences that exist between English and Japanese RCs (Schachter, 1974). Thus, the specific measure that we have chosen for the present study 'relative clauses per AS-unit' not only reflects subordination but it can be classified as a cognitively difficult measure related to Japanese B2 level learners. In terms of 'relative clauses per AS-unit' being a reliable tool for L2 development, as our targeted RC types consist of simple (OS) and complex (OPREP) forms, our measure is able to capture both forms as it represents all RC types. Finally, subordination is a suitable measure of syntactic complexity for the purpose of the present study due to the intermediate proficiency level of the participants. Norris & Ortega (2009) point out that coordination measures are more suitable for beginner level learners as syntactic complexity first occurs through the use of coordinated clauses. Subordination however, becomes the more dominant form of syntactic complexity at the intermediate stage of L2 development, whilst complexity at the phrasal level, for example, nominalisation is mainly achieved with advanced learners. Consequently, subordinate measures appear to be a suitable indicator of syntactic complexity for the intermediate participants of this study. Specifically, the use of 'relative clauses per AS-unit', although not perfect, does appear to measure a grammatical feature known for its cognitive difficulty with Japanese learners and it therefore reflects our definition of syntactic complexity.

The next sub-section outlines the measure of the final dependent variable for oral development of the main study: fluency.

5.7.4 Fluency

For the purpose of this study, fluency was defined in 2.4.1 as "the rapid, smooth, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (adapted from Lennon, 2000, p. 26). As this definition comprises of different aspects of fluency we will briefly review them again in order to clarify the measure chosen for the present study:

1. 'Rapid' concerns the speed of L2 delivery i.e. the ability to produce speech in real-time speaking conditions that do not involve conscious planning time (Ellis & Barkhuizen, 2005).
2. 'Smooth' relates to the use of formulaic language as mentioned in 2.3.2 which allows the learner to produce 'chunks' of language which are easier and faster to produce as opposed to processing individual units of language. Thus smooth also involves 'rapid' speech.
3. 'Lucid' relates to the ability to produce L2 speech which is understandable to others. In terms of this study, lucid speech refers to pronunciation and intonation. Other factors such as grammatical accuracy are considered as a separate measure for accuracy, as outlined in 5.7.2.
4. 'Efficient translation of thought or communicative intention into language' refers to the process of L2 speech production outlined in Kormos' (2011) model in 2.3.1. This involves the ability to process conceptualisation, formulation and articulation efficiently in order to produce L2 speech without time delays.
5. 'temporal constraints of on-line processing' refers to the ability to produce language under the natural time constraints of everyday speech which typically does not involve planning time.

Although we acknowledged the limitations of this definition in terms of additional aspects of oral fluency that may not be accounted for, it does appear to cover all the main areas that were to be investigated in the present study. Given this definition, an appropriate measure was required. In the pilot study, two speech rate measures were taken from Levkina & Gilabert (2010) and Gilabert (2007a): pruned (rate B) and unpruned (rate A) speech rates. The former involved 'the total number of syllables produced per minute excluding repetitions, self-corrections, false-starts, L1 use and

incomprehensible language' whilst the latter consisted of 'the total number of syllables produced per minute'. However, as the main study involved a larger sample size that would involve more data analysis, only one measure for fluency could be used. As a result, the 'pruned speech rate' measure was chosen which is defined as "the average number of syllables produced per minute of pruned speech, i.e. speech from which repetitions, false starts and other performance features have been excluded" (Levkina & Gilabert, 2010, p. 182). This measure was convenient for the purpose of this study as it enabled us to measure fluency according to our definition in the following way:

1. Syllables per minute is a speech rate measure and therefore represents point one of our definition 'rapid' as we can measure how fast learners produce syllables. This measure also represents point two: 'smooth' as it can calculate the amount of formulaic language produced by counting the syllables of 'chunked language'.
2. Excluding "repetitions, false starts and other performance features" (Levkina & Gilabert, 2010, p. 182) enables us to measure language in relation to point four: 'efficient translation of thought or communicative intention into language' as we are only focusing on L2 language intended by the speaker, thus discarding unnecessary language related to repetitions and false starts. Other performance features concerns self-corrections, L1 use and incomprehensible language which reflects point three: 'lucid' L2 language that is understandable.

Finally, in terms of point five: 'under the temporal constraints of on-line processing' fluency was measured under testing conditions that did not involve planning time. In other words, the pre- and post-test narratives did not involve strategic planning so the learners were required to engage in 'on-line processing'. Consequently, the 'pruned speech rate' measure 'the total number of syllables produced per minute excluding repetitions, self-corrections, false-starts, L1 use, incomprehensible language' used in our testing conditions with oral narratives that did not involve strategic planning time enabled us to measure fluency according to our definition. Following Gilabert (2007a), the formula for calculating pruned speech remained the same as in the pilot study:

$$\frac{\text{Total number of syllables (excluding repetitions, self-corrections, false starts, L1 use and incomprehensible language)}}{\text{Total number of seconds}} \times 60$$

As discussed earlier in 5.3.1, to prevent the use of strategic planning during the pre- and post-tests, the present study adopted a similar stance to the no-planning conditions of Yuan & Ellis (2003) in which the participants were allowed to briefly look at the narrative for approximately twenty seconds prior to performance. As a result, using the calculation above would enable us to see how much language was produced under real-time speaking conditions during the pre- and post-tests. Finally, as in the pilot study (4.4.3), the conditions for measuring fluency remained the same, for example, in terms of total speaking time, syllables were recorded from the participant's use of *'Today…'* and finished at the end of the narration. Phrases such as *'that's all'*, *'finished'* were excluded. For instructions of how fluency was coded and analysed using the computer software CLAN, as well as further examples of what constituted repetitions, self-corrections, false-starts and L1 use, see appendix U.

Although the pruned speech rate measure appears to be a suitable match in terms of our definition of fluency, there are still limitations regarding the use of speech rate measures in general. For example, 'syllables per minute' could prove problematic as an indication of L2 proficiency because syllables could include L1 use. Furthermore, a learner could repeatedly use the same words again and again thus sounding incoherent yet would appear to be fluent due to the amount of syllables produced. However, the steps taken to exclude language in this study's analysis of pruned speech appears to counter the above limitations aimed at speech rate measures as the present study does not include L1 use thus the measure can serve as an indication of L2 proficiency. Furthermore, as the measure also does not include repetitions, a learner can only appear fluent by producing coherent speech. Finally, although pruned speech appears to be a more suitable match in terms of our definition, it was also considered to be a more reliable measure than unpruned speech in terms of assessing a learners' ability to communicate effectively in the L2. For example, "speech rate B has the advantage of eliminating the meaningless speech (e.g repetitions) which may be used by L2 speakers to gain time and to give the impression that they are being fluent" (Levkina & Gilbert, 2012, p. 182). Thus it allows us to focus on the learners' intended L2 output, as mentioned in point four of our definition. In addition, pruned speech has also been used as a measure in previous studies (Mehnert, 1998, Ortega, 1999, Gilabert, 2007b). To conclude, this measure may still be limited as we know that fluency is a complex phenomenon which appears impossible to measure in every way. However, the pruned

speech rate measure chosen for this study does allow us to measure fluency according to our definition.

Having discussed all the measures of our dependent variables for oral L2 development, the next sub-section describes the analysis of the grammatical judgement test.

5.7.5 Grammatical judgement test

In line with Izumi (2003), the analysis of the grammatical judgement test involved removing any items which the participant did not answer or did not follow the instructions as outlined in 5.3.2 (for example, not providing a written answer to an incorrect item). As this created an unequal amount of test items for the participants as a whole, each learner's grade was determined by "dividing the total correct scores by the total number of applicable items" (p. 301). Consequently, each participant was given a percentage score for the pre- and post-tests. The distracter items were not included in the analysis, and in addition, if a learner corrected an item which was considered ungrammatical but was not related to the targeted RC errors or use of the accompanying 3rd person singular, the item was also discarded from the analysis. RC related corrections would involve rectifying the RC errors discussed in 5.3.2 For example:

- *The woman who you met her went to the hospital (pronoun retention)*
  Correction by deleting a pronoun: *The woman who you met went to the hospital*

- *The woman is young who likes John (nonadjacency)*
  Moving a relative marker to the correct position next to the head noun: *The woman who is young likes John*

- *I looked for the book who Tom was talking about (incorrect relative marker morphology)*
  Correcting a relative marker: *I looked for the book which Tom was talking about*

- *The girl was in pain saw the dentist (inappropriate relative marker omission)*

Correction by inserting a relative marker: *The girl who was in pain saw the dentist*

- *He wants the car which have big wheels (incorrect verb tense)*
  Correcting verb tense in relation to the head noun: *He wants the car which has big wheels*

In line with Izumi (2003), any non-RC related corrections such as change of prepositions or articles, the item would be considered non-applicable. In terms of the 3rd person singular morphology that accompanied the targeted RC types, typical error corrections would involve:

- *Peter believe that he played a piano which was made in France'. (incorrect subject-verb agreement)*
  Correct subject-verb agreement: *Peter believes/believed that he played a piano which was made in France.*

If however, the participant corrected an error relating to 3rd person singular but then made a mistake with the connecting RC type, the item would be scored incorrect. The items were scored by the researcher with 1 point for a correct answer and O points for an incorrect answer. Finally, in line with Izumi (2003) and Reinders & Cho (2012), the items of the grammatical judgement tests were checked for reliability using the computer software Kuder Richardson 20 (KR-20) which measures the consistency of the items within each test. The results are listed below in comparison with one of the results from Reinders & Cho (2012):

Table 28. *Grammatical judgement test reliability values*

	Pre-test	Immediate post-test	Delayed post-test
Present study (GJT)	0.80	0.69	0.88
Reinders & Cho (2012) (GJT)	0.850	0.619	N/A

KR values of .80 or more suggest high levels of reliability concerning the measure (Izumi, 2003). Reinders & Cho (2012) reported high levels of reliability for their test with the exception of the post-test (0.619). The present study's test also exhibited high values apart from the immediate post-test (0.69) thus overall the values reflect high reliability levels regarding the participants' scores.

So far, we have described the quantitative data analysis measures relating to research question one. We now turn to the qualitative data analysis measures that relate to research question two: namely the post-task interviews.

5.7.6 Qualitative data analysis: post-task interviews

Following Ortega (2005), the qualitative data concerning the GP and GUP group's planning strategies involved a "content analysis of emergent themes" (p. 83) in order to identify the areas of planning that each group focused on as they prepared for narratives that increased in complexity. According to Dornyei (2007) content analysis of qualitative research "follows the very generalized sequence of coding for themes, looking for patterns, making interpretations, and building theory" (p. 46). In the case of task planning, possible patterns that could emerge from the post-task interview data could be attention towards grammar. Any patterns that emerged from the learners' metacognitive responses about the cognitive processes they engaged in during planning would help us to make interpretations about *how* guided planning and task complexity contributed towards L2 oral development. However, Dornyei (2007) points out that interpretations based on qualitative data analysis are problematic due to the issue of subjectivity. For example, subjective data analysis may result in the researcher using data to support preconceived beliefs of what he/she hopes to find rather than analysing data objectively to see what it *actually* reveals. Issues relating to subjectivity can be reduced in different ways. Ortega's (2005) study which investigated learners' planning strategies did not contain any research questions or hypotheses. In doing so, Samuda & Bygate (2008) commented that this approach enabled Ortega (2005) to analysis the data impartially and the findings could be considered less subjective. In terms of the present study, hypotheses concerning the GP and GUP group's planning strategies were omitted to help counter the problem of subjectivity. In doing so, the finding of this study can be considered less impartial due to the lack of preconceived beliefs regarding the learners'

planning strategies. Subjectivity was further countered by the use of computer-aided qualitative data analysis (CAQDAS) which will be discussed later in this sub-section.

The content analysis was carried out in three phases. The first phase involved obtaining the most frequently used words from the participants' responses to the main interview questions listed in 5.6.2. This was achieved by transcribing the interview data from questions one to six into the software program CLAN and then running a word frequency program which provided an output file containing the frequency of each word used by each participant. To identify words associated with planning, only nouns and verbs were targeted whilst pronouns and articles such as 'I' and 'the' were excluded from the analysis. Robson (2002) notes that caution must be taken with word frequency lists as words can be used with different meanings therefore it is important to return to the data set to confirm how the word was used. As a result, the second phase of the analysis involved identifying example phrases containing the most commonly used words from each participant, and categorizing them in an excel document. Following Ortega (2005), the third phase of the analysis was to identify common themes from the examples provided which represented the participants' planning strategy such as *'focus on grammar'* or *'focus on story'*.

The analysis also investigated whether the learners incorporated additional planning strategies which were targeted in question four *(did you think about grammar, vocabulary, how to organise the story, or something else?)*. The researcher noted any additional strategies in a separate column of the excel document. The planning strategies of each participant were then categorized together to determine the overall planning strategies of each group. A further frequency word count was then carried out which examined the most commonly used words within each group as a whole in order to confirm whether the most frequently used words reflected the common theme of each group's planning strategy. This procedure was initially carried out with the transcriptions taken at the start of the task treatment in week two. The process was then repeated with the transcriptions taken from weeks three and four to determine each group's planning strategies during each week of the treatment. However, weeks three and four included an additional analysis of question five which examined possible changes in the learners' planning strategies from the previous week *(did you plan differently compared to last week?)*. The researcher then compared the overall strategies of each group from weeks two to four for any patterns or differences as each group attempted more complex tasks over time.

Dornyei (2007) outlines the advantages of CAQDAS. Qualitative research usually involves collecting large sums of data which is time consuming to transcribe and results in pages and pages of documents that can be difficult to sort through in order to retrieve key information. CAQDAS is invaluable in this regard as computers can save significant amounts of data in electronic files which are quick and easy to access. Furthermore, "content analysis programs can search for and count key domain-specific words and phrases" (p. 264) such as word frequency lists that can instantly provide the most commonly used words from a data set. In doing so, CAQDAS provides researchers with objective evidence to support their argument as it reduces the bias and subjectivity criticisms levelled at qualitative analysis. In addition, it saves the researcher a lot of time and effort attempting to carry out the same functions manually. CAQDAS is also particularly beneficial with mixed methods research as qualitative information can be easily linked to quantitative software enabling triangulation of data by comparing quantitative and qualitative information of participants. There are however certain disadvantages when using CAQDAS in that electronic files are susceptible to viruses whilst data is also at risk of being accidently removed or deleted. CAQDAS also requires training so researchers can operate software programs for analysis, and professional assistance can be expensive and difficult to access.

Although a number of CAQDAS software programs are available, such as Nvivo, the present study opted for the use of CLAN and the CHILDES program. As outlined above, CLAN could perform word frequency lists which provided the most commonly used words of the GP and GUP learners. This analysis provided a level of objectivity in the interpretations about what aspects of planning learners focused on that was omitting from the analysis in the pilot study. CLAN was also useful in terms of triangulation of data and the mixed methods design of this study as both quantitative coding relating to CAF and qualitative data was analyzed and stored within the same software. In addition, the CHILDES Program was free to download and it included a user-friendly manual which explained how to use it. The reliability of the software was assured as it had been used in previous planning studies (for example, Levkina & Gilbert, 2010, Tavakoli & Foster, 2008). Finally, all copies of the data were saved onto hard-drives and portable USB's which cancelled out the threat of data loss.

5.8 Chapter conclusion

This chapter has described the research methodology for the main study and justified each method used. The amendments made to the main study were:

1. A larger sample size of 13 learners per group. The control group was removed and the main study involved the comparison of two different planning groups: 'guided planning'(GP) , and 'guided and unguided planning' (GUP).
2. The 'unguided' planning condition of the pilot study was changed to 'guided and unguided' planning in order to draw learners' attention towards the targeted forms at the start of the treatment and to see whether learners could independently plan towards the forms as they progressed with tasks that increased in complexity over time.
3. The simple RC type OO was replaced by the more complex OPREP RC type in order to provide a greater linguistic challenge for the B2 participants. In addition, cognitive state verbs were replaced by the use of 3rd person singular and plural as the targeted forms of the study.
4. A grammatical judgement test was included as an alternative assessment to test whether the effects of guided planning and task complexity, as well as guided and unguided planning and task complexity produced developmental gains of the targeted forms.
5. A rating scale was used to measure learners' accuracy of the targeted forms instead of 'error-free' measures. In addition, only one measure was used for syntactic complexity and fluency: 'relative clauses per AS-unit' and 'pruned speech rate' respectively.
6. The planning time of both group's treatment was gradually reduced as they attempted more complex tasks in order to prime learners to perform the post-tests under no planning conditions.
7. The hypotheses for research question two were removed in order to reduce the subjectivity of the data analysis and findings concerning the learners' planning strategies.
8. CAQDAS software (CLAN) was used to analyse the qualitative post-task interview data for emerging themes and to produce frequency word lists.

9. A questionnaire was included to provide information relating to the equality of both groups in terms of their preferences towards communication and accuracy when speaking in the L2 prior to the task planning treatment.

Finally, features of the main study that remained the same were as follows:

1. The simple OS RC type remained a targeted form of the study.
2. The pre- immediate and delayed post-test design and the duration of the study remained the same: 7 weeks.
3. Treatment tasks were sequenced with an increase in complexity along intentional reasoning demands.
4. The oral proficiency of the participants remained the same at B2 (CEFR) level.
5. Oral narratives remained the same task-type for oral assessment and task treatment.
6. The post-task interview questions remained largely unchanged.

Having discussed all the changes made to the methodology of the main study, we are now ready to move onto the results of the study. The following chapter displays and discusses all the quantitative results relating to research question one whilst chapter seven provides the qualitative results for research question two.

6. QUANTITATIVE RESULTS

6.1 Introduction

The analysis of the present study is split into two chapters. This chapter will discuss the quantitative data related to each of the four hypotheses concerning research question one, whilst the following chapter will analyse the qualitative data related to the two hypotheses for research question two. This chapter begins in 6.2 by providing the descriptive statistics for each of the three dependant variables used to measure oral development (fluency, accuracy and complexity) as well as the additional dependent variable: the grammatical judgement test used to measure the learners' receptive awareness of the targeted forms. Sub-sections 6.2.1 – 6.2.4 then describes each of the four hypotheses relating research question one:

- To what extent does guided planning and task complexity facilitate L2 oral development in terms of fluency, morphological accuracy of OS and OPREP English relative clauses and 3rd person singular and plural, as well as syntactic complexity for second year Japanese university learners of English?

- Hypothesis one: guided planning and task complexity leads to less developmental gains in fluency compared to guided and unguided planning and task complexity.

- Hypothesis two: guided planning and task complexity leads to greater developmental gains in morphological accuracy of OS and OPREP English relative clauses and 3rd person singular and plural compared to guided and unguided planning and task complexity.

- Hypothesis three: guided planning and task complexity leads to greater developmental gains in syntactic complexity compared to guided and unguided planning and task complexity.

- Hypothesis four: guided planning and task complexity leads to greater developmental gains in learners' receptive awareness of OS and OPREP English relative clauses and 3rd person singular and plural compared to guided and unguided planning and task complexity.

6.2 Descriptive statistics of the dependent variables

The first step of data analysis for the present study was to provide the descriptive statistics relating to all the dependent variables to determine whether outliers existed within the sample data (Dornyei, 2007, Field, 2009). Outliers relate to any values that are incompatible with the rest of the data set. They often consist of extreme figures, greater or lesser than the rest of the sample, and as a result, they create non-normal distribution of values within a data set which distorts the overall result and creates an inaccurate representation of the sample (Dornyei, 2007). Field (2009) points out that large sample sizes of more than 30 tend to exhibit normal distribution regardless. As this data set was using a total sample size of n = 26, frequency tests were carried out to test for normal distribution for each dependent variable. Field (2009) illustrates how frequency tests can show whether a data sample is normally distributed or not by analysing the degree of skewness and kurtosis. The larger the values of skewness and kurtosis, the greater the degree of non-normality within a data set. Skewness and kurtosis values that are zero indicate normal distribution whilst values that are close to one or more (positive or negative) indicate non-normal distribution (see table 29).

Table 29. *Descriptive statistics of the dependent variables measured at the pre- and post-tests (n-size = 26): means, standard deviations, skewness and kurtosis*

Dependent Variable	Mean	Std. Deviation	Skewness	Kurtosis
Fluency: pruned rate B (pre-test)	35.57	15.00	1.15	1.97
Fluency: pruned rate B (immediate post-test)	58.46	15.66	1.54	4.34
Fluency: pruned rate B (delayed post-test)	59.32	13.95	.57	-.23
Complexity: relative clauses per AS-unit (pre-test)	0.07	0.18	2.73	6.67
Complexity: relative clauses per AS-unit (immediate post-test)	0.73	0.21	-0.530	0.659
Complexity: relative clauses per AS-unit (delayed post-test)	0.77	0.25	0.24	3.30
Accuracy: rating scale (pre-test)	7.00	0.47	2.63	6.988
Accuracy: rating scale (immediate post-test)	37.65	2.29	-.386	-.111
Accuracy: rating scale (delayed post-test)	40.58	2.66	-.008	.826
Grammatical judgement pre-test	68.87	13.82	-.18	.04
Grammatical judgement immediate post-test	85.26	9.45	-1.64	3.21
Grammatical judgement Delayed post-test	83.55	9.98	-1.47	2.82

As can be seen from table 29, over half of the dependent variables measured at the pre- and post-tests obtain values greater than one, for example, the fluency rate B pre-test measure shows skewness of 1.15, whilst the complexity AS-unit pre-test measure shows skewness of 2.73. Consequently, these results imply that a number of outliers are present in the data set and that non-normal distribution exists. Field (2009) reports the consequences of data that contain non-normal distribution. Many statistical analyses such as General linear models carried out on software programs such as SPSS

operate on the assumption that data is normally distributed. If this is not the case i.e. the sample data is non-normal, then most statistical analyses such as ANCOVAs are prohibited as they will not provide accurate statistics. However, steps can be taken to identify and correct or remove outliers from data sets so that samples can be transformed towards normal distribution. Following Field (2009), Shapiro-Wilks tests were carried out to provide a visual display of the outliers present in each of the dependent variables. Once outliers are identified, they can be removed from data sets and further tests can be re-run to test for normal distribution (Dornyei, 2007). Field also points out that "if our analysis involves comparing groups, then what's important is not the overall distribution but the distribution in each group" (2009, p. 147). As this study involves the comparison of two planning groups, Shapiro-Wilk tests were also used to test for normality within each group for each dependent variable. A significance value of $p < 0.05$ suggests that non-normal distribution exists within each group. The results of the Shapiro-Wilk tests are shown in table 30.

Table 30. *Shapiro-Wilk tests of normal distribution within each group per dependent variable*

Tests of Normality			
	Group	Shapiro-Wilk	
		Df	Sig.
Fluency: rate B pre-test	GP	12	0.547
	GUP	12	0.680
Fluency: rate B immediate post-test	GP	12	0.788
	GUP	12	0.449
Fluency: rate B delayed post-test	GP	12	0.320
	GUP	12	0.131
Grammatical judgement pre-test	GP	13	0.109
	GUP	13	0.602
Grammatical judgement immediate post-test	GP	13	0.083
	GUP	13	0.280
Grammatical judgement delayed post-test	GP	13	0.730
	GUP	13	0.154
Complexity: relative clauses per AS-unit pre-test	GP	13	0.000
	GUP	13	0.000
Complexity: relative clauses per AS-unit immediate post-test	GP	13	0.003
	GUP	13	0.211
Complexity: relative clauses per AS-unit delayed post-test	GP	13	0.002
	GUP	13	0.912
Accuracy: rating scale pre-test	GP	13	0.001
	GUP	13	0.000
Accuracy: rating scale immediate post-test	GP	13	0.324
	GUP	13	0.545
Accuracy: rating scale delayed post-test	GP	13	0.815
	GUP	13	0.607

Note. GP = Guided planning and task complexity, GUP = guided and unguided planning and task complexity.

As can be seen in table 30, removing outliers for certain variables allows for normal distribution. For example, in terms of fluency, participant 109 from the GP group, and participant 217 from the GUP group, both exhibited large fluency scores (80.03 and 45.52 syllables per minute respectively in the pre-test compared to overall sample mean of 35.57 as shown in table 29). As a result, these values skewed the overall outcome of each group. However, when these participants were removed (n = 12 per group) both groups exhibited normal distribution which is reflected in a significance values of $p > 0.05$ at the pre- and post-tests as displayed in table 30. In terms of the grammatical judgement tests, it was not necessary to remove outliers because the pre- and post-tests displayed normality within each group with $p > 0.05$. However, the complexity AS-unit measure showed non-normality within each group with values of $p < 0.05$. In addition, the accuracy rating scale measure showed a mixture of normal and non-normal values. For example, non-normality exists at the accuracy pre-test measure for both groups ($p < 0.05$) whilst normality exists at the immediate post-test for the GP group ($p = .324$) and the GUP group ($p = .545$).

Although outliers were identified within the complexity and accuracy variables, unlike fluency, removing one outlier from each group did not result in normal distribution. In terms of accuracy, only two evident outliers were present at the pre-test, one outlier from the GP group: participant 105 and one outlier from the GUP group: participant 214 who both scored 14 and 15 respectively compared to the overall mean average of 7 as displayed in table 29). Despite removing these outliers, deviations in the GUP group's pre-test results were still significant. It was therefore decided that both 105 and 214 would remain in their respective groups for two reasons. According to Field (2009) removing an outlier, "should only be done if you have good reason to believe that this case is not from the population that you intend to sample" (p. 153). Participants 105 and 214 were considered part of the intended overall sample, the only difference was that they produced two relative clauses each during the pre-test whilst the rest of the sample did not. As the narratives were designed to elicit seven contexts of the RCs, their exclusion for producing the form hardly seems justified. In addition, due to both groups' small sample size (n = 13), removing more than one outlier would impact negatively on the outcome of the study. As we saw in 4.4.5, the pilot study group sizes were n = 2 which prevented inferential statistical analysis from taking place. Furthermore, as we discussed in 5.2.1, the GP and GUP group sizes were already reduced from n = 17 to n = 13 in order to include only those participants who had taken

the placement test. Consequently, to main a healthy sample size per group, the complexity and accuracy variables were kept at n = 13 but the results concerning these variables were limited as a result of non-normal distribution.

Despite the limitation of the complexity and accuracy sample sets, the focus of this study was to compare two groups that were considered equal in order to evaluate two types of planning and task sequencing conditions. As a result, additional tests were carried out on the accuracy and complexity measures to ensure there were no significance differences *between* the two groups prior to engaging in their respective task sequencing treatments. To do this, independent sample *t*-tests were performed on the accuracy and complexity pre-test scores to compare the means of both groups to see whether there were any significant differences between them. In terms of the accuracy pre-test, the GP group obtained a higher mean score (M = 7.15, SD = 2.30) compared to the GUP group (M = 6.85, SD = 2.54) however the result of the *t*-test showed this difference was not significant $t(24) = .323$, $p > 0.05$. Both groups could therefore be considered roughly equal at the pre-test in terms of accuracy. Regarding complexity, after removing two outliers (participant 105 from GP group and participant 218 from GUP group) there was no significant differences between the groups $t(22) = .434$, $p > 0.05$, thus the GP and GUP groups could also be considered equal at the pre-test regarding syntactic complexity. Finally, to confirm the extent to which both groups could be considered equal at the pre-test, further independent samples *t*-tests were carried out on the two remaining dependent variables: fluency and the grammatical judgement test as well as the placement test scores. The results showed no significant differences between both groups' pre-test scores in terms of fluency: GP (M = 28.65), GUP (M = 37.96) $t(17.253) = -1.985$, $p > 0.05$, the grammatical judgement test: GP (M = 72.18), GUP (M = 65.55) $t(24) = 1.235$, $p > 0.05$, and the placement test which we saw earlier in 5.2.1: GP (M = 561.38), GUP (M = 544.62) $t(24) = 1.262$, $p > 0.05$. These results therefore suggest that both groups could be considered equal in terms of the dependent variables of the study prior to the task sequencing treatment. Thus, both groups provided a level platform that would enable us to see which task sequencing treatment would be more effective in terms of L2 oral development.

So far we have examined the GP and GUP's normality of distribution scores to see whether both groups could be considered equal in terms of the dependent measures used. Frequency tests confirmed outliers were present in the data for some the dependent variables. However, as this study involved the comparison of two groups,

Shapiro-Wilk tests showed that normal distribution existed within each group for fluency at n = 12, and the grammatical judgement test at n = 13. Non-normality existed within each group for accuracy and complexity however as the study's sample size was too small to remove numerous outliers, the accuracy and complexity samples remained non-normal. To ensure both groups were considered roughly equal at the pre-test, independent sample *t*-tests showed there were no significant differences between the two groups with accuracy at n = 13 and complexity at n = 12. Consequently, both groups were considered equal at the pre-test, and this was also the case in terms of fluency, the grammatical judgement test and the placement test scores. We can now turn to analyse which task treatment, guided planning and task complexity, or guided and unguided planning and task complexity was more effective in terms of L2 oral development. We now begin our analysis of hypotheses one to four starting with fluency.

6.2.1 Results of hypothesis one: effects on fluency

According to hypothesis one, guided planning and task complexity leads to less developmental gains in fluency compared to guided and unguided planning and task complexity. Hypothesis one was not confirmed. The descriptive results taken from the paired samples *t*-test for the fluency 'pruned speech rate B' measure relates to syllables per minute excluding repetitions, self-corrections, false starts, L1 usage and incomprehensible language are displayed in table 31. It also contains the mean of the GP and GUP group's pre- and post-test narrative scores, their standard deviations, the mean differences of the pre- and immediate post-test, the pre- and delayed post-test, and the immediate and delayed post-test.

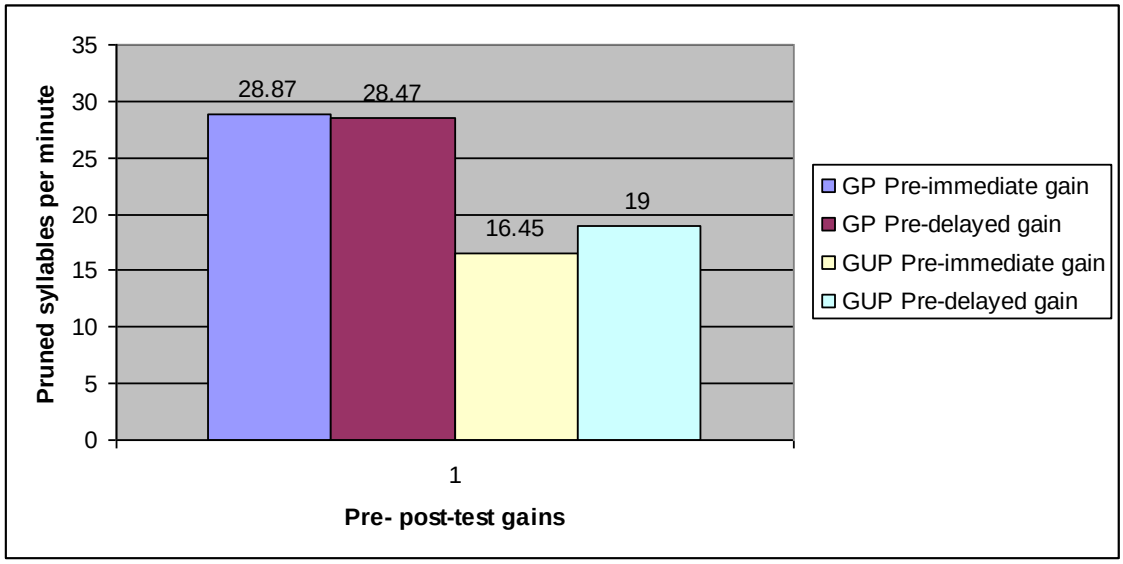
Table 31. *Descriptive statistics for fluency: GP and GUP groups*

	Rate B pruned speech rate (excluding repetitions, self-corrections, false starts, L1 use and incomprehensible language					
Group (n=12)	Pre-test	Immediate Post-test	Delayed Post-test	Pre-immediate post-test difference	Pre-delayed post-test difference	Immediate delayed post-test difference
GP Mean	28.65	57.51	57.12	28.87	28.47	-0.40
SD	7.93	13.20	12.01	14.70	13.36	6.13
GUP Mean	37.96	54.41	56.96	16.45	19.00	2.55
SD	14.20	10.18	12.27	14.39	8.40	14.70

Note. GP = guided planning group, GUP = guided and unguided planning group

At the pre-test, we can see that the GUP group produced more syllables per minute (*M* = 37.96) than the GP group (*M* = 28.65). However, GP group produced a larger mean gain from pre-test to the immediate post-test (*M* = 28.87) compared to the GUP group (*M* = 16.45). The GP mean gain remained largely unchanged from the pre-test to the delayed post-test as the difference was only -0.4 syllables per minute between both post-tests. The GUP group's mean improved from the pre- to the delayed post-test by 2.55 syllables per minute although the pre- delayed gain (*M* = 19) was still less than the GP group (*M* = 28.47). Figure 13 provides a visual display of the results.

Figure 13. *GP and GUP group's fluency pre- and post-test gain scores*



To determine the statistical significance of the mean gains within each group and the extent of their effect sizes, paired sample *t*-tests and Cohen's *d*-values were calculated (see table 32).

Table 32. *Statistical significance and effect sizes for fluency: GP and GUP group*

Group	n-size	Pre-immediate post-test (Sig.)	Pre-immediate post-test Cohen's *d*-value	Pre-delayed post-test (Sig.)	Pre-delayed post-test Cohen's *d*-value	Immediate delayed post-test (Sig.)
GP	12	.000	2.77	.000	2.92	0.827
GUP	12	.000	1.39	.000	1.61	0.560

These results support the descriptive statistics as they display significant differences from the pre-test to the immediate post-test for the GP group $t(11) = -6.80$, $p < 0.05$, and the GUP group $t(11) = 3.96$, $p < 0.05$. The effect sizes are also large for both groups ($d > 0.8$). Similar results are shown from the pre-test to the delayed post-test for the GP group $t(11) = -7.38$, $p < 0.05$, and the GUP group $t(11) = 7.83$, $p < 0.05$, with large effect sizes again ($d > 0.8$). Finally, there were no significant differences from the immediate to the delayed post-test for the GP group $t(11) = .22$, $p > 0.05$ and the GUP group $t(11) = .60$, $p > 0.05$ which implies that the effects of both treatments remained stable over time.

To see whether the pre- post-test mean gains *between* the two groups were statistically significant, independent sample *t*-tests and Cohen's *d*-values were calculated (see table 33).

Table 33. *Statistical significance and effect sizes between the GP group and the GUP group's pre- and post-test mean gains for fluency*

	Pruned speech rate B (excluding repetitions, self-corrections, false starts, L1 use and incomprehensible language)		
	Pre- immediate post-test	Pre- delayed post-test	Immediate delayed post-test
Sig.	0.048	0.049	0.531
Cohen's *d* value	0.89	0.89	0.2

In terms of gains from the pre-test to the immediate post-test, there was a significant difference between the GP group and the GUP group $t(22) = 2.09$, $p < 0.05$. This was supported by a large effect ($d = 0.89$) between both groups. In terms of gains from the pre-test to the delayed post-test, there was also a significant difference between both groups $t(22) = 2.08$, $p < 0.05$ with a large effect size for fluency between the two groups ($d = 0.89$). Finally, there were no significant differences between both groups' immediate and delayed post-test gain scores $t(14.713) = .53$, $p > 0.05$. In summary, the pre- immediate post-test results showed that guided planning and task complexity produced significantly greater gains in fluency compared to guided and unguided planning and task complexity. These results also remained stable from the pre- to the delayed post-test which suggests that guided planning and task complexity produced significant long-term gains in fluency compared to guided and unguided planning and task complexity.

6.2.2 Results of hypothesis two: effects on accuracy

According to hypothesis two, guided planning and task complexity leads to greater improvements in morphological accuracy of OS and OPREP English relative clauses, as well as 3rd person singular and plural compared to guided and unguided planning and task complexity. Hypothesis two was confirmed.

As mentioned in 5.7.2, accurate production of the targeted forms was measured using a nine point rating scale. The maximum score permitted in producing the seven obligatory cases of RC's and the accompanying use of 3rd person singular and plural in each pre- and post-tests was 60. The results are displayed in table 34. It displays the mean of the GP and GUP group's pre- and post-test narrative scores, their standard deviations, and the mean gain differences of the pre- and immediate post-test, the pre- and delayed post-test, and the immediate and delayed post-test.

Table 34. *Descriptive statistics of the GP and GUP groups for gains in morphological accuracy*

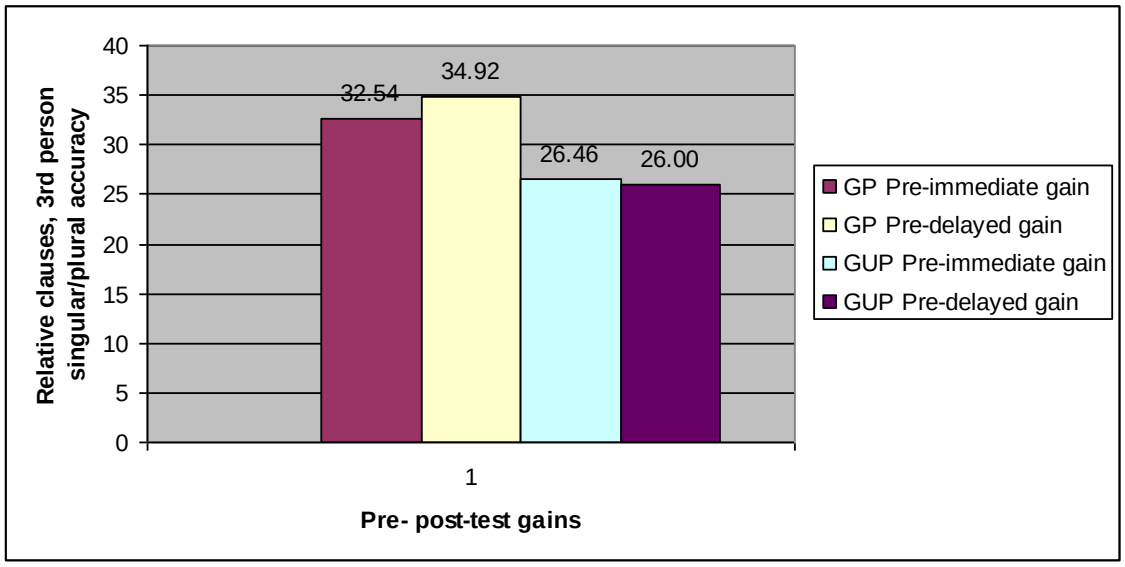
	Morphological accuracy rating scale					
Group (n=13)	Pre-test	Immediate Post-test	Delayed Post-test	Pre-immediate post-test difference	Pre-delayed post-test difference	Immediate delayed post-test difference
GP Mean	7.15	39.69	42.08	32.54	34.92	2.38
SD	2.30	6.69	6.81	6.97	7.12	6.92
GUP Mean	6.85	33.31	32.85	26.46	26	-0.46
SD	2.54	11.92	11.97	11.84	12.52	7.81

Note. GP = guided planning group, GUP = guided and unguided planning group

Both the GP and the GUP groups produced low means at the pre-test narrative (M = 7.15 and M = 6.85) respectively. The GP group produced a larger mean gain from the pre-test to the immediate post-test (M = 32.54) compared to the GUP group (M = 26.46). The GP group produced an even larger mean gain from the pre-test to the delayed post-test with a difference of (M = 2.38) between both post-tests. The GUP

group, on the other hand, produced a smaller mean gain from the pre-test to the delayed post-test with a difference of (M = -0.46) between both post-tests (see figure 14).

Figure 14. *GP and GUP group's pre- and post-test gain scores in morphological accuracy*



The statistical significance of the pre- and post-test mean scores within each group and their effect sizes are shown in table 35.

Table 35. *Statistical significance and effect sizes of the GP and GUP group's pre- and post-test gains in morphological accuracy*

Group	n-size	Pre-immediate post-test (Sig.)	Pre-immediate post-test Cohen's *d*-value	Pre-delayed post-test (Sig.)	Pre-delayed post-test Cohen's *d*-value	Immediate delayed post-test (Sig.)
GP	13	.000	6.77	.000	7.15	0.238
GUP	13	.000	3.20	.000	3.13	0.835

These results display significant differences from the pre-test to the immediate post-test for the GP group $t(12) = 16.83$, $p < 0.05$ and the GUP group $t(12) = 8.06$, $p < 0.05$. There were also large effect sizes for both groups ($d > 0.8$). Similar results are shown from the pre-test to the delayed post-test for the GP group $t(12) = 17.68$, $p < 0.05$, and the GUP group $t(12) = 7.49$, $p < 0.05$. Once again both groups produced large effect

sizes (d > 0.8). Finally, there were no significant differences from the immediate to the delayed post-test for the GP group $t(12) = -1.24$, $p > 0.05$, and the GUP group $t(12) = 2.13$, $p > 0.05$ which implied that the effects of both treatments remained constant over time.

To account for the significant gains in accuracy of the GP and the GUP group, an additional analysis was carried out to examine the extent to which both groups produced the targeted forms (see table 36). Table 36 displays the obligatory contexts of the targeted forms for each test, the maximum target-like performance score for each test, the GP and GUP's mean average for each test and their percentage correlation to target-like performance. On the basis of the low pre-test scores for both groups, it appears the pre-test narrative was not successful in eliciting the targeted RC types and 3rd person singular and plural. However, after both groups' respective treatment, the GP group and the GUP group were able to produce accurate instances of the targeted forms which were closer to target-like performance, particularly at the delayed post-test.

Table 36. *The GP and GUP group's pre- and post-test comparison to target-like performance*

Pre-post Tests	Obligatory contexts of RCs	Target-like performance	GP	Percentage	GUP	Percentage
Pre-test	7	60	7.15	11.92%	6.85	11.42%
Immediate Post-test	7	60	39.69	66.15%	33.31	55.52%
Delayed Post-test	7	60	42.08	70.13%	32.85	54.75%

The effects of guided planning and task complexity helped produce high means for the GP group. For example, the delayed post-test (M = 42.08) is over 70% accurate in terms of target-like performance concerning the seven obligatory contexts of the targeted forms. The effects of guided and unguided planning and task complexity helped produce high means for the GUP group although not as high as the GP group. For example, the GUP group achieved 55.52% accuracy in terms of target-like performance at the immediate post-test, and 54.75% accuracy at the delayed post-test.

Further analysis was also carried out to determine whether the GP and the GUP learners produced the expected RC types (OS and OPREP) at the seven obligatory contexts during each narrative test. As discussed in 5.3.1, each narrative contained five instances of OS and two instances of OPREP. Table 37 shows the number of participants from the GP and the GUP groups who produced the expected RC types, the number of participants who produced an alternative RC type, for example, producing OS instead of OPREP, the number of participants who described the content of the picture without producing the form (simplified). Finally, those learners who avoided describing the content all together, in other words, they did not describe the picture that contained an obligatory RC context.

Table 37. *The GP and GUP group's pre- and post-test production of the OS and OPREP RC types*

Pre- and post-tests	RC production							
Pre-post tests	Avoided	%	Simplified	%	Alternative RC	%	Expected RC	%
Pre-test (GP)	11	12.09	75	82.42	2	2.2	3	3.3
Pre-test (GUP)	10	10.99	79	86.81	2	2.2	0	0
Immediate post-test (GP)	0	0	12	13.19	12	13.19	67	73.63
Immediate post-test (GUP)	0	0	23	25.27	15	16.48	53	58.24
Delayed post-test (GP)	0	0	11	12.09	6	6.59	74	81.32
Delayed post-test (GUP)	0	0	24	26.37	13	14.29	54	59.34

Note. GP = guided planning group, GUP = guided and unguided planning group

The total number of expected RC's for each group was 91 (seven contexts per narrative with 13 learners per group). At the pre-test, we see that only 3 expected RC types were produced by the GP group whilst the GUP group produced none and only a further 2 alternative RC types were produced by each group. The vast majority of both groups described the contexts without using any relativization whilst a smaller amount did not describe the contexts at all. Thus, the GP and GUP's group's low pre-test scores in accuracy were as a result of describing the narrative without using relativization. A different story however is depicted at the immediate post-test where the GP group produced 73.63% of the 91 expected RCs. The GUP group did not produce as many of

the expected RC types (58.24%) as 16.48% of the participants used alternative RC types whilst 25.27% described the contexts without using relativization. In terms of the delayed post-test, the GP group produced 81.32% of the 91 expected RC types. The GUP group however, only produced 59% of the expected RC types. These results therefore show the strong effect that guided planning and task complexity had on the GP group's production of the targeted RC types OS and OPREP compared to guided and unguided planning and task complexity. We now turn to see whether the pre- post-test mean gains *between* the two groups was statistically significant (see table 38).

Table 38. *Statistical significance and effect sizes between the GP group and the GUP group's pre- and post-test mean gains for morphological accuracy*

	Morphological accuracy of relative clauses and 3rd person singular or plural		
	Pre- immediate post-test	Pre- delayed Post-test	Immediate delayed post-test
Sig.	0.124	0.035	0.335
Cohen's *d* value	0.65	0.91	0.40

There were no statistically significant differences between the GP and the GUP group in terms of gains from the pre-test to the immediate post-test $t(24) = 1.60$, $p > 0.05$. This was supported by a medium effect between the two groups ($d = 0.65$) in terms of accuracy development. Differences between the two groups were significant however from the pre-test to the delayed post-test $t(24) = 2.23$, $p < 0.05$ whilst ($d = 0.91$) suggests a large effect size for morphological accuracy between the two groups. Finally, there were no significant differences between both groups' immediate and delayed post-test gain scores $t(24) = 0.98$, $p > 0.05$. In summary, the pre- immediate post-test results showed that guided planning and task complexity produced greater gains in morphological accuracy of the targeted forms compared to guided and unguided planning and task complexity. However, the pre- to the delayed post-test results showed that guided planning and task complexity produced significant long-term gains in morphological accuracy compared to guided and unguided planning and task complexity.

6.2.3 Results of hypothesis three: effects on syntactic complexity

According to hypothesis three, guided planning and task complexity leads to greater gains in syntactic complexity compared to guided and unguided planning and task complexity. Hypothesis three was partially confirmed.

The descriptive results taken from the syntactic complexity measure (relative clauses per AS-unit) are displayed in table 39. It illustrates the mean average of the GP and GUP group's pre- post-test narrative scores, their standard deviations, and the mean differences of the pre- and immediate post-test, the pre- and delayed post-test, and the immediate and delayed post-test.

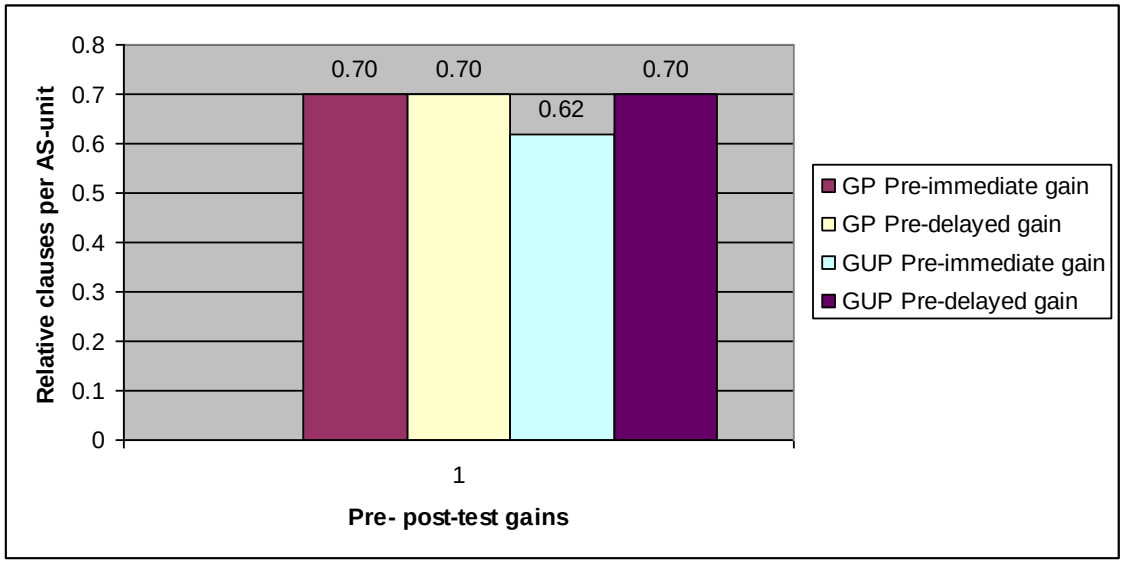
Table 39. *Descriptive statistics of the GP and GUP groups for syntactic complexity*

	Relative clauses per AS-unit					
Group (n=12)	Pre-test	Immediate Post-test	Delayed post-test	Pre-immediate post-test difference	Pre-delayed post-test difference	Immediate-delayed post-test difference
GP Mean	0.06	0.76	0.76	0.70	0.70	0
SD	0.17	0.13	0.06	0.22	0.19	0.18
GUP Mean	0.04	0.66	0.74	0.62	0.70	0.08
SD	0.10	0.26	0.35	0.25	0.35	0.26

Note. GP = guided planning group, GUP = guided and unguided planning group

Both the GP and the GUP groups produced low means at the pre-test narrative (M = 0.06 and M = 0.04 respectively) in terms of relative clauses per AS-Unit. However, the GP group produced a larger mean gain from the pre-test to the immediate post-test (M = 0.70) compared to the GUP group (M = 0.62). The GP group produced the same mean gain at the delayed post-test with a difference of 0 between both post-tests. The GUP group's mean improved by 0.08 at the delayed post-test thus both groups produced the same mean gain (M = 0.70) in terms of relative clauses per AS-unit (see figure 15).

Figure 15. *GP and GUP group's syntactic complexity pre- and post-test gains*



The statistical significance of the pre- and post-test mean scores within each group and their effect sizes are shown in table 40.

Table 40. *Statistical significance and effect sizes of the GP and GUP groups for syntactic complexity*

Group	n-size	Pre-test immediate post-test (Sig.)	Pre-test immediate post-test Cohen's *d*-value	Pre-test delayed post-test (Sig.)	Pre-test delayed post-test Cohen's *d*-value	Immediate delayed post-test (Sig.)
GP	12	.000	4.83	.000	5.74	0.943
GUP	12	.000	3.29	.000	2.84	0.286

These results show that the mean gains from the pre-test to the immediate post-test were significant for the GP group $t(11) = -10.85$, $p < 0.05$ and the GUP group $t(11) = -8.76$, $p < 0.05$. The effect sizes were also large for both groups ($d > 0.8$). Similar results are shown from the pre-test to the delayed post-test for the GP group $t(11) = -12.92$, $p < 0.05$ and the GUP group $t(11) = -6.88$, $p < 0.05$ with large effect sizes in the amount of relativization for both groups ($d > 0.8$). Finally, there were no significant differences from the immediate post-test to the delayed post-test for the GP group $t(11) = 0.74$, $p > 0.05$ and the GUP group $t(11) = -1.121$, $p > 0.05$ which confirms that the effects of both treatments remained constant over time. These results support the similar

findings for accuracy in the previous sub-section. To see whether gains *between* the two groups were statistically significant, let us examine table 41.

Table 41. *Statistical significance and effect sizes between the GP group and the GUP group's pre- and post-test mean gains for syntactic complexity*

	Relative clauses per AS-unit		
	Pre-test immediate Post-test	Pre-test delayed post-test	Immediate post-test delayed post-test
Sig.	0.417	0.955	0.31
Cohen's *d* value	0.35	0	0.37

There were no significant differences between the GP group and the GUP group in terms of gains from the pre- to the immediate post-test $t(22) = .83$, $p > 0.05$. In addition, ($d = 0.35$) indicated a small to medium effect for relativization between the two groups. Differences between the groups were even less regarding the pre-test to the delayed post-test gains $t(22) = -0.06$, $p > 0.05$, whilst ($d = 0$) implied there was no effect between the groups. Finally, there were no significant differences between both groups' post-test gain scores $t(15.40) = 1.05$, $p > 0.05$. In summary, the pre- immediate post-test results showed that guided planning and task complexity produced greater gains in syntactic complexity compared to guided and unguided planning and task complexity. Although the GP group's results remained stable from the pre- to the delayed post-test, the GUP group's results increased slightly which confirms that guided planning and task complexity produced similar long-term gains in syntactic complexity compared to guided and unguided planning and task complexity.

6.2.4 Results of hypothesis four: effects on receptive awareness of the targeted forms

According to hypothesis four, guided planning and task complexity leads to greater gains in learners' receptive awareness of OS and OPREP RC types and 3rd person singular and plural compared to guided and unguided planning and task complexity. Hypothesis four was partially confirmed.

As discussed in 5.7.5, each participant received a percentage value for each pre- and post grammatical judgement test that was determined by dividing the total number of correct items by the total number of completed items. The descriptive results are displayed in table 42. It illustrates the mean average of the GP and GUP group's pre- and post-test narrative scores, their standard deviations, and the mean differences of the pre- and immediate post-test, the pre- and delayed post-test, and the immediate and delayed post-test.

Table 42. *Descriptive statistics of the GP and GUP group's grammatical judgement test*

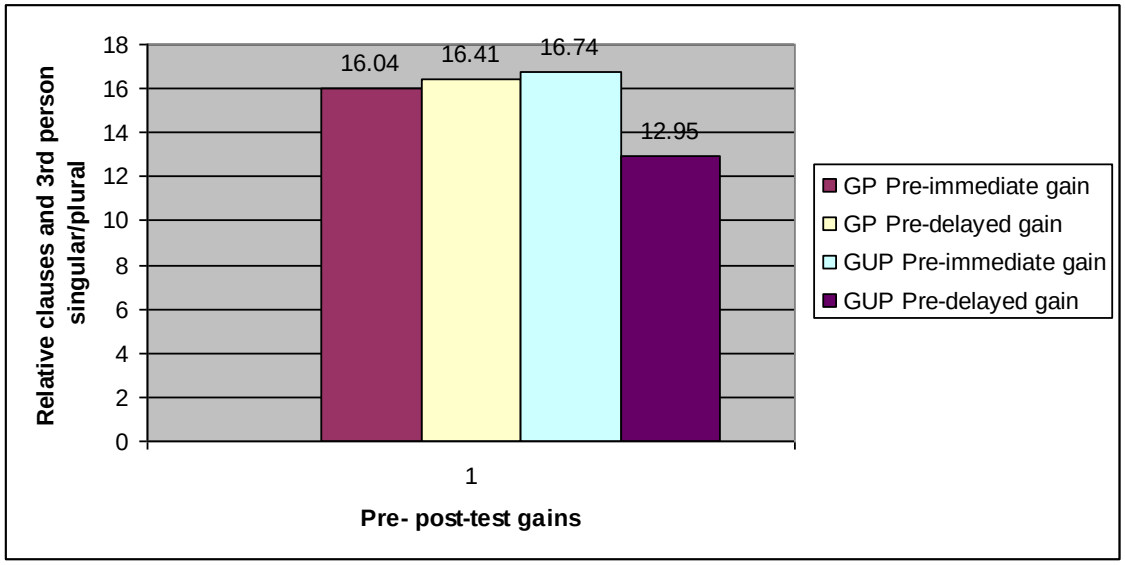
	Grammatical judgement test					
Group (n = 13)	Pre-test	Immediate Post-test	Delayed post-test	Pre-immediate post-test difference	Pre-delayed post-test difference	Immediate delayed post-test difference
GP Mean	72.18	88.22	88.59	16.04	16.41	0.37
SD	12.83	4.48	5.13	10.06	10.79	4.44
GUP Mean	65.55	82.29	78.50	16.74	12.95	-3.79
SD	14.48	12.12	11.23	15.74	12.02	8.38

Note. GP = guided planning, GUP = guided and unguided planning

At the pre-test, the GP group scored a higher mean average (*M* = 72.18%) compared to the GUP group (*M* = 65.55%). Although the GP group produced a slightly smaller mean gain from the pre-test to the immediate post-test (*M* = 16.04%) compared to the GUP group (*M* = 16.74%). The GP group produced a similar mean gain at the

delayed post-test with a difference of 0.37% between both post-tests. However, the GUP group's mean dropped by 3.79% at the delayed post-test (see figure 16).

Figure 16. *GP and GUP grammatical judgement test pre- and post-test gains*



The statistical significance of the pre- and post-test mean differences for the GP and GUP groups and their effect sizes are shown in table 43.

Table 43. *Statistical significance and effect sizes of the GP and GUP group's pre- and post grammaticality judgement test scores*

Group	n-size	Pre-immediate post-test (Sig.)	Pre-immediate post-test Cohen's *d*-value	Pre-delayed post-test (Sig.)	Pre-delayed post-test Cohen's *d*-value	Immediate delayed post-test (Sig.)
GP	13	0.000	1.74	0.000	1.75	0.771
GUP	13	0.002	1.30	0.002	1.04	0.129

These results support the descriptive statistics as they display significant differences from the pre-test to the immediate post-test for the GP group $t(12) = -5.75$, $p < 0.05$ and the GUP group $t(12) = 3.84$, $p < 0.05$. The effect sizes were also large for receptive awareness of the targeted forms for both groups ($d > 0.8$). Similar results are reported for both groups from the pre-test to the delayed post-test as the GP group showed significant gains $t(12) = -5.48$, $p < 0.05$, as did the GUP group $t(12) = 3.89$, $p <$

0.05 with large effect sizes for both groups (d > 0.8). There were no significant differences from the immediate post-test to the delayed post-test for the GP group $t(12) = -.298$, $p > 0.05$ and the GUP group $t(12) = -1.63$, $p > 0.05$ which confirms that both groups' treatment had a lasting effect over time. We now look at the differences between the two groups in table 44.

Table 44. *Statistical significance and effect sizes between the GP and GUP group's pre- and post grammatical judgement tests*

	Grammatical judgement test		
	Pre- immediate Post-test	Pre- delayed post-test	Immediate Delayed post-test
Sig.	0.894	0.448	0.131
Cohen's *d* value	0.06	0.32	0.53

There were no significant differences between the GP group and the GUP group in terms of gains from the pre-test to the immediate post-test $t(24) = -.135$, $p > 0.05$ and this consequently showed a very small effect for receptive awareness of the targeted forms between the two groups ($d = 0.06$). Differences between both groups were also not significant from the pre-test to the delayed post-test $t(24) = .772$, $p > 0.05$ whilst ($d = 0.32$) confirms a small to medium effect between both groups. Finally, differences between both groups' immediate and delayed post-test gains were not significant $t(18.24) = 1.58$, $p > 0.05$. In summary, the pre- immediate post-test results showed that guided planning and task complexity produced a smaller mean gain in learners' receptive awareness of the targeted forms compared to guided and unguided planning and task complexity. However, the results of the pre- delayed post-tests showed that guided planning and task complexity produced greater long-term gains in learners' receptive awareness of the targeted forms compared to guided and unguided planning and task complexity but this difference was not statically significant.

6.3 Chapter conclusion

We began this chapter by investigating the normality distribution of the data samples for each of the dependent measures. The grammaticality judgement test produced normal distribution at n = 13, the fluency measure produced normality after reducing the group sample size to n = 12. The accuracy and complexity values however remained non-normal so to reduce this limitation, independent sample *t*-tests were carried out which confirmed that both groups had roughly equal scores at the pre-test in terms of accuracy (n = 13) and complexity (n = 12). Thus overall, both groups were considered equal in terms of the dependent measures used and the results of hypotheses one to four are summarized as follows:

- Hypothesis one: not confirmed. The GP group produced significant gains in fluency compared to the GUP group.

- Hypothesis two: confirmed. The GP group produced significant gains in morphological accuracy of OS and OPREP RC types and accompanying use of 3rd person singular and plural compared to the GUP group.

- Hypothesis three: partially confirmed. The GP group produced greater gains in syntactic complexity from the pre- to the immediate post-test compared to the GUP group but this gain was not significantly different whilst both groups produced the same gain from the pre- to the delayed post-test.

- Hypothesis four: partially confirmed. The GP group produced a smaller gain from the pre- to the immediate post-test but a larger mean gain from the pre- to the delayed post-test compared to the GUP group however these differences were not statistically significant.

In chapter eight, we discuss the findings and implications of these results in detail.

7. QUALITATIVE RESULTS

7.1 Introduction

The purpose of this chapter is to report the qualitative results of this study's second research question:

- What strategies do Japanese second year university learners of English use when planning for oral narratives that increase in complexity over time?

Before we examine the results of this question, 7.2 first reviews the findings of the pilot study from chapter four. 7.2.1 then provides a brief review of the treatment tasks and planning conditions of the GP and GUP groups whilst 7.2.2 reviews the interview questions used to elicit the learners' planning strategies. 7.3 then analyses the GP group's planning strategies during the task sequencing treatment. 7.4 then summarises the results. In 7.5 the same analysis is presented for the GUP group, and finally 7.6 summarises the results.

7.2 Pilot study review

The pilot study results which were discussed in 4.6 showed that the B1 and B2 learners who received guided planning initially attended to either grammar or the storyline but then focused on grammar only as they prepared for more complex tasks over time. The B1 and B2 unguided planners initially planned their narratives by focusing on grammar, and as they attempted more complex tasks, their strategies remained largely unchanged over time. Consequently, the results showed that this sample of B1 and B2 learners generally focused on grammar when preparing for oral narratives, whether under the influence of guided planning or unguided planning.

In terms of the present study, the GP and GUP groups were interviewed about their planning strategies immediately after they had performed an oral narrative in weeks two, three and four of the task sequencing treatment. Before we begin our analysis, let us first remind ourselves of the task sequence both groups performed, their planning conditions and the interview questions used.

7.2.1 The treatment tasks

As mentioned in 5.4, five narratives were designed for the treatment task sequence. They were all different in terms of storyline and they increased in cognitive complexity through increasing intentional reasoning demands by containing additional cases of RCs (see table 45).

Table 45. *Obligatory cases of RCs per treatment narrative*

	Task treatment: Increasing intentional reasoning demands		
	Narrative 1 Week 2	Narrative 2 & 3 Week 3	Narrative 4 & 5 Week 4
Relative Clauses	7	9	10

The treatment tasks further increased in complexity by increasing resource-dispersing demands which as discussed in 5.4.1 involved reducing planning time. The first treatment task (narrative one) involved ten minutes guided planning time for both groups which consisted of grammar guidance notes and examples of the targeted RC types OPREP and OS, and the accompanying use of 3rd person singular. Narratives two and three however consisted of seven minutes planning time whilst narratives four and five involved four minutes planning time. The GP group continued to receive guided planning throughout the task sequence however the GUP group were not allocated grammar guidance from task two onwards and were instead instructed to plan independently (see table 46).

Table 46. *Strategic planning conditions of the GP and GUP groups during the task sequencing treatment*

Groups	Narrative 1	Narrative 2 & 3	Narrative 4 & 5
GP	Guided planning	Guided planning	Guided planning
GUP	Guided planning	Unguided planning	Unguided planning

Note. Length of planning time = narrative 1 (10 minutes), narratives 2 & 3 (7 minutes), narratives 4 & 5 (4 minutes).

7.2.2 Interview questions

As discussed in 5.6.2, in order to elicit metacognitive responses about the cognitive processes that the participants were engaged in during strategic planning, the following six main questions were used:

Q1 Did you plan?
Q2 How did you plan?
Q3 What was your focus when you prepared the story?
Q4 Did you think about grammar, vocabulary, how to organize your story, or something else?

During weeks three and four, the following questions were used in order to compare planning strategies with the previous week:

Q5 Did you plan differently compared to last time?
Q6 In what way?

Let us now turn to the interview analysis concerning the planning strategies of the GP group.

7.3 Results of the GP group's planning strategies

We begin our analysis by examining the results of the GP group's planning strategies at the start of the treatment in week two which are displayed in table 47. It outlines the identification number of each participant, the most frequently used word per participant, an example phrase containing the word, additional planning strategies that were prompted by question four ('N/A' implies no additional strategy was mentioned). Finally, an interpretation of each learner's planning strategy is provided. Two learners' transcripts of the interview data can be found in appendix V.

Table 47. *Week two: GP group's planning strategies*

GP group	Frequently used word	Frequency	Example	Additional Planning Strategy (Question 4)	Overall Planning Focus
101	Story	6	"umm, think the story"	"umm clear sentence"	Story and grammar
102	Grammar	3	"I focus on the grammar"	N/A	Grammar
103	Grammar	3	"umm, looking the picture and use this grammar"	N/A	Grammar
104	Long	3	"long ears and long hairs"	"grammar….umm, oh, story"	Grammar and story
105	Easy	4	"easy style, easy writing"	"vocabulary…ah grammar"	Vocabulary and grammar
106	Grammar	2	"I focus grammar"	N/A	Grammar and vocabulary
107	Grammar	3	"ah, umm, grammar"	N/A	Grammar
108	Grammar	2	"grammar and story"	"grammar and story"	Grammar and story
109	Grammar	2	"ahh, focus? Ehh, to use the grammar"	N/A	Grammar
110	Grammar	3	"umm, I am careful about special grammar"	"grammar, vocabulary"	Grammar and vocabulary
111	Thinking	4	"umm, umm, my thinking, my thinking, I wrote my thinking"	"umm stories and this steps and grammars"	Story and grammar
112	Grammar	2	"using this grammar"	"umm grammar…and story"	Grammar and story
113	Grammar	2	"yes, grammar, story, vocabulary"	"yes, grammar, story, vocabulary"	Grammar, story, vocabulary

Let us first examine participant 101's strategies before moving onto the group as a whole. The most frequently used word by 101 is *'story'* which was produced six times. An example phrase used by 101 which contained the word was *"Umm, think the story"* which indicates that 101's planning strategy involved thinking about the story of the narrative. When asked to provide an example of planning for the story, 101 responded *"which rabbits they want"* which reflects the pet-shop storyline of narrative one. As the word *'story'* was the most frequently used word by 101 suggests that the learner's main planning strategy was focusing on the story, as opposed to thinking about vocabulary or grammar. Although in response to whether 101 focused on anything else (question four) 101 replied *"umm clear sentence"* which could imply that 101 was focusing on grammar in order to speak accurately. As a result, we can assume that 101's planning strategy was targeted towards conceptualizing the story and focusing on grammar in order to speak clearly.

The rest of the GP group provided different responses and frequently used words. Participants 102, 103, 106, 107 and 109 did not indicate an additional planning strategy hence they were categorized 'N/A'. From the above table, we can see that the most frequently used word from most of the participants when describing their planning strategies was *'grammar'*. After reviewing the example phrases provided by each participant using their most frequently used word, the common theme of the GP group for describing their planning strategies appears to be *'using grammar'*. This assumption was further supported by the results of the most frequently used words of the GP group as a whole during week two's post-task interviews (see table 48). As can be seen, the word *'grammar'* was used the most by the GP group at 26 occurrences (see appendix W for the complete list of frequency words used by the GP group).

Table 48. *Most frequently used words of the GP group during week two*

Word	Frequency
Grammar	26
Story	12
This	11
And	11
Which	10
Example	9

All 13 participants either mentioned a planning focus towards grammar during questions two or three, or mentioned grammar as an additional planning strategy when prompted in question four. Six GP participants focused on the story as well as grammar. The participants' interpretation of the story was in line with the storyline of narrative one, for example, "*Ah, ahh character of rabbits", "Lisa went to the pet shop with her parents".* Four participants attended towards vocabulary as well as grammar during guided planning whilst participant 113 focused on all three aspects of planning.

If we compare the GP group's planning strategies at week two with the results of the pre-study questionnaire on L2 communication and accuracy in 5.6.3, we saw that seven learners (54%) of the GP group expressed an interest in communication over accuracy when speaking in the L2. However, 13 of the participants focused on grammar during guided planning. Following Ortega (2005), if we are to assume that a focus on form is synonymous with accuracy, then learners who have an L2 oral orientation towards communication appear willing to focus on form and accuracy when provided with grammar guidance during planning. We can therefore conclude that the GP group's main planning strategy during week two was to use the targeted grammar forms, whilst approximately half the participants attended to additional aspects of planning such as conceptualizing the story and focusing on vocabulary.

7.3.1 Week three: GP group's planning strategies

The planning strategies for the GP group during week three are displayed in table 49. Following the analysis of week two, table 49 displays the participant's identification number, the most frequently used word (N/A implies no word was frequently used), an example phrase, additional planning strategies (N/A implies there was no additional strategy). It also includes responses to question five which examined whether the learners' planning strategy had changed from the previous week (N/A implies the learner did not respond). Finally, an interpretation of the learner's planning strategy is provided.

Table 49. *Week 3: GP group's planning strategies*

GP group	Frequently used word	Frequency	Example	Additional planning strategy? (Question 4)	Different from last week? (Question 5)	Overall planning Focus
101	N/A	N/A	"umm, which car they want"	"umm, grammar"	"no"	Grammar
102	Thinks	2	"umm, I, I use thinks or wants"	N/A	"no"	Grammar
103	Which	2	"umm, how to use which, which"	N/A	"same"	Grammar
104	Story	2	"ahh, yes, story and vocabulary"	"umm, this grammars"	"no"	Story, vocabulary and grammar
105	Has	2	"for example, the car which has, has"	N/A	"umm, no"	Grammar
106	Story	2	"umm, I think story, story"	N/A	"umm, not different"	Story
107	Recline	2	"recline, I recline the amount of writing"	"grammar"	"little"	Grammar
108	grammar	2	"umm, special grammar"	"umm, focus this picture"	"same"	Grammar, story
109	grammar	3	"ah I was focused on grammar"	N/A	"yes"	Grammar
110	Is	5	"erm, character is many brother, character is"	"grammar and vocabulary"	"same"	Grammar and vocabulary
111	N/A	N/A	"grammar"	"vocabulary"	N/A	Grammar, vocabulary
112	focused, explaining	2	"I focused on explaining the, umm, umm, I focused on the explaining the car"	"all of them"	"no"	Grammar, vocabulary, Story
113	Which	2	"ehh, which has"	N/A	"umm, no"	Grammar

At week three, there are slightly less frequently used words from most of the participants compared to week two, for example participant 101 and 111 did not repeat the same word more than once from the question responses that were analyzed. This could be due to the fact that question two, *"how did you plan?"* was only used on a few students as the researcher relied more on question three, *"what did you focus on when you were planning?"* which was similar in meaning and elicited greater responses from the participants. As a result, by week three, question two was not relied upon as much to elicit the GP and GUP group's planning strategies. On the other hand, the omission of question two should have been compensated by the use of question five which was used as an additional question to determine whether learners' strategies had changed from the previous week. Consequently, there should have been the same opportunity to respond in the L2, yet there was still a drop in the most frequently used words from the GP group.

As in week two, the most frequently used words from most of the GP participants when describing their planning strategies was *'grammar'* and *'which'*. Both of these words can be considered synonymous as the latter was the targeted RC marker. After reviewing the example phrases provided by each participant, the common theme of the GP group appeared to be *'using grammar' or 'using which'.* Our interpretation was supported by the results of the most frequently used words of the GP group as a whole during week three's post-task interviews (see table 50). The word *'which'* and *'grammar'* was used the most by the GP group at 16 and 13 occurrences respectively.

Table 50. *Most frequently used words of the GP group during week three*

Word	Frequency
Which	16
Grammar	13
This	9
Is	9
Yes	8
Use	8

All the participants bar 106 either mentioned a planning focus towards grammar during questions two or three, or mentioned grammar as an additional planning strategy

when prompted in question four. One participant (111) attended to vocabulary as well as grammar during guided planning, one participant (108) focused on the story as well as grammar, and two participants (104, 112) focused on all three aspects of planning. Finally, all the participants except 109 confirmed that they used the same planning strategies in week three as they did in week two. Most of the GP group commented on their continual attention towards the targeted grammar forms, for example, *"umm, using which", "umm, use this sentence", "when the picture is bubble I use thinks or believe".* Only participant 109 said that he planned differently compared to week two and this appeared to be on account of the different storylines and characters in the tasks, for example, *"three brothers, no three brothers".*

We can therefore conclude that the GP group's main planning strategy during week three was to use the targeted grammar forms, whilst a small number of participants attended to additional aspects of planning such as conceptualizing the story as well as vocabulary. Finally, most of the GP group confirmed that they used the same planning strategies as in week two.

7.3.2 Week four: GP group's planning strategies

The planning strategies of the GP group during week four are displayed in table 51. It displays the participant's identification number, the most frequently used word (N/A implies no word was frequently used), an example phrase, additional planning strategies (N/A implies there was no additional strategy). It also includes responses to question five which examined whether the learners' planning strategy had changed from the previous week. Finally, an interpretation of the learner's planning strategy is provided.

Table 51. *Week 4: GP group's planning strategies*

GP group	Frequently used word	Frequency	Example	Additional Planning Strategy? (Question 4)	Different from last week? (Question 5)	Overall Planning Focus
101	N/A	N/A	"grammar"	N/A	"same"	Grammar
102	Example, Sentences	2	"for example, red light or blue light"	N/A	"the same"	Grammar
103	Which	3	"umm, if is which have, etto kocchi, which has, umm"	N/A	"same"	Grammar
104	Wants, Plural	3	"she think ehh she wants she thinks she wants she thinks she wants"	N/A	"same"	Grammar
105	Example, Picture	2	"for example, sarah believes that she likes the car"	N/A	"erm yes"	Grammar
106	Think	3	"umm, I think pro, ah, I'm look, I'm looking for the proper word in English, I think Japanese"	"ahh vocabulary"	"ahh, iya, umm, almost same"	Vocabulary and grammar
107	Use	3	"ahh, umm, I ahh, I'm used to using grammar so umm, I, umm, I tell you the story, umm better"	N/A	"umm, no"	Grammar and story
108	Grammar	2	"Erm grammar is he want the cat which the teachers is talking about"	N/A	"ah different"	Grammar and story
109	This	4	"er for this for this picture, the girl which which for this grammar ah for this picture first I describe sec second sister and next I describe the third sister"	N/A	"did you plan er yes"	Grammar
110	Which	2	"Erm er I'm careful about which has or which have"	"grammar"	"no same"	Grammar
111	Same	2	"contents is the same"	N/A	"same"	Grammar
112	Using	2	"I focused on using grammar"	N/A	"erm almost the same"	Grammar
113	Story	3	"umm, story and ehh, detail"	N/A	"ahh same"	Grammar and story

At week four of the treatment, there was a slight increase in the most frequently used words compared to week three, although participant 101 did not repeat the same word more than once from the questions that were analyzed. Question four was only used on a few students who did not provide a response to question three. For example, if the students answered question three i.e. *'what did you focus on when you were planning?'* question four (*did you think about anything else, for example, grammar or vocabulary?)* was considered unnecessary at this stage of the treatment because the participants were by now accustomed to the interview process and were responding about their strategies without the need for prompting.

At week four, there was more variety in the most frequently used from the GP group however the most commonly used words across the group appeared to be similar to weeks two and three i.e. *'which', 'use', 'example'*. Again, these words can be associated with the examples of the targeted RC grammar guidance provided. After reviewing the example phrases provided by each participant, the common theme of the GP group again appears to be use of the targeted grammar for example, *'using grammar' or 'using which have/has'.* Once again, the common theme was supported by the results of the most frequently used words of the GP group as a whole during week four's post-task interviews (see table 52). The word *'which'* and *'grammar'* was used the most by the GP group at 14 and 12 occurrences respectively.

Table 52. *Most frequently used words of the GP group during week four*

Word	Frequency
Which	14
Grammar	12
For	11
Same	10
Use	8
Think	8

All 13 participants either mentioned that their planning focus was directed towards grammar during question three, or that their planning strategy was similar to the previous week which also focused on grammar. Three participants (107, 108, 113) attended to the story as well as grammar during guided planning, and one participant

(106) focused on vocabulary as well as grammar. Finally, most of the participants except for 105, 108 and 109 confirmed that they used the same planning strategies in week four as they did in week three, for example, "*erm er I'm careful about which has or which have", "umm bubbles she thinks thinks and wants the car which umm".* There is also evidence at this stage of the treatment that some GP participants had started to become accustomed to preparing and using the grammar forms, for example, *"you give me example sentences so etto (umm) I can, nan do ka, (sometimes) I can see the sentences so it is easy it is easily for me to write".* Other participants however, (105, 108 and 109) said that they planned differently compared to week three. 105 said his planning strategy changed because *"I think this picture become more difficult....for example erm picture two cars er".* In other words, the increase in task complexity along intentional reasoning demands by having more content within the pictures, in this case two cars instead of one, resulted in the planning conditions being more linguistically challenging. In addition, 108's planning strategy changed because*"I have to write a short time".* In this case, the increase in task complexity along resource-dispersing dimensions through the reduction of planning time prevented the learner from planning in as much detail as in week three. Finally, 109 said *"the ac erm actor was different before"* implying that his planning strategy had changed because the narrative contained different characters compared to the previous week.

We can therefore conclude that the GP group's main planning strategy during week four was use of the targeted grammar forms, whilst a few participants attended to additional aspects of planning such as the story and vocabulary. Finally, most of the GP group confirmed that they used the same planning strategies as in week three.

7.4 Conclusion of the GP group's planning strategies

Despite the results of the GP group's pre-study questionnaire in which half the participants expressed a preference towards oral communication over accuracy, the above analysis showed that the planning strategies of the GP group during weeks two, three and four appeared to remain largely focused towards grammar. Initially, at week two, the most frequently used word from most of the participants was *'grammar'* and the common theme of the group's planning strategies was *'using grammar'.* At week three, as the tasks increased in complexity, the majority of learners still used the words

'grammar' and *'which'* the most during their interviews, and the common theme again was *'using grammar'* or *'using which'*. Finally, at week four, their planning strategies appeared the same as the common theme was *'using grammar'* or *'using which have/has'*. Most the GP group also confirmed that their planning strategies remained largely unchanged as the tasks increased in complexity during week two, three and four. Consequently, these results confirmed that the GP group initially focused on grammar and then maintained their attention on form as they prepared for more complex narratives over time. In other words, their planning strategies remained unchanged throughout the treatment.

7.5 Results of the GUP group's planning strategies

The analysis of the GUP group's planning strategies was carried out in the same format as the GP group. We begin our analysis by examining the results of the GUP group's planning strategies at week two which are displayed in table 53. It shows the identification number of the participant, the most commonly used word per participant, an example phrase containing the word, additional planning strategies that were prompted by question four (N/A implies no additional strategy was mentioned). Finally, an interpretation of the learner's planning strategy is provided.

Table 53. *Week two: GUP group's planning strategies*

GUP Group	Frequently used word	Frequency	Example	Additional planning strategy? (Question 4)	Overall planning focus
214	grammar, story	2	"umm umm to tell story umm grammar is important"	"ah, and story"	Grammar and story
215	Grammar	2	"I think, umm, it's very important to tell, Umm, umm, I mean using grammar"	N/A	Grammar
216	Grammar	3	"grammar and story"	"grammar and story"	Grammar and story
217	Grammar	5	"focus, ah, before we practice grammar Styles, I think, ah, using the grammar"	N/A	Grammar
218	Grammar	4	"grammar is good"	"story, grammar, vocabulary"	Story, grammar, Vocabulary
219	Focus	5	"focus? Focus? I focus, focus on the grammar"	N/A	Grammar
220	Grammar	4	"ah, gamm, grammar"	N/A	Grammar
221	Is	5	"in this picture, a human is her mother and, She, the rabbit which she wants is long ear"	N/A	Grammar
222	story	2	"yeah, mainly, thinks story"	No	Story
223	Who	2	"who and what who did"	"yeah mainly thinks story"	Story, grammar, Vocabulary
224	Imagine	3	"I didn't, ehh, image, imagine, image, imagine, interesting imagine, I didn't make interesting story"	"umm grammar"	Grammar
225	Sentence	4	"I planned the sentence, when I, umm, think, I saw that rabbits which have long, which has long hair"	"grammar"	Grammar
226	N/A	N/A	"they want the rabbits. Mother thinks that she likes the rabbit which has long ears"	"grammar….umm story"	Grammar and story

As can be seen from table 53, the most frequently used word from most of the GUP participants when describing their planning strategies was *'grammar'*. After reviewing the example phrases provided by each participant, the common theme of the GUP group appeared to be *'using / focusing on grammar'*. This assumption was supported by the results of the most frequently used words of the GUP group as a whole during week two's post-task interviews (see table 54). The word *'grammar'* was used the most by the GUP group at 33 occurrences (see appendix W for the complete list of frequency words used by the GUP group).

Table 54. *Most frequently used words of the GUP group during week two*

Word	Frequency
grammar	33
This	15
Is	13
Story	11
Focus	11
To	10

All bar one of the participants (222) either mentioned a planning focus towards grammar during questions two or three, or mentioned grammar as an additional planning strategy when prompted in question four. Three participants (214, 216, 226) attended to the story as well as grammar during guided planning, for example, *"thinking the situation…umm if I go to pet shop what do I think?"*, *"ehh, talking a the picture"*. Two participants (218, 223) focused on all three aspects of planning whilst one participant (222) focused on the story only.

If we compare the GUP group's planning strategies at week two with the results of the pre-study questionnaire on L2 communication and accuracy in 5.6.3, we can see that 6 learners (46%) of the GUP group expressed an interest towards communication over accuracy when speaking in the L2, yet 12 of the participants focused on grammar during guided planning. These results are similar to the GP group in that certain GUP learners whose L2 oral orientation was towards communication appeared to have focused more on form and accuracy when provided with grammar guidance during guided planning. We can therefore conclude that the GUP group's main planning

strategy during week two was to use the targeted grammar forms, whilst a small number of participants incorporated other aspects of planning such as thinking about the story and the required vocabulary.

7.5.1 Week three: GUP group's planning strategies

The planning strategies of the GUP group during week three are displayed in table 55. It displays the participant's identification number, the most frequently used word, an example phrase, additional planning strategies (N/A implies there was no additional strategy). It also includes responses to question five which examined whether the learners' planning strategy had changed from the previous week. Finally, an interpretation of the learner's planning strategy is provided.

Table 55. *Week three: GUP group's planning strategies*

GUP group	Word	Frequency	Example	Additional planning strategy? (Question 4)	Different from last Week?(Question 5)	Overall Planning Focus
214	Think	2	"ahh, I think similar"	N/A	"ahh, I think similar"	Story and Grammar
215	Grammar	2	"umm, I learn grammar"	N/A	"this is easier"	Grammar and Story
216	Construction	2	"umm, umm construction"	"story"	"same same"	Story and Grammar
217	Doll	3	"what the doll is wear, or wearing"	N/A	"umm, yes"	Grammar and vocabulary
218	Which	2	"for example, simon wants the car which have red tyre"	"story line"	"I forget last week"	Grammar and Story
219	Grammar	3	"umm, I focused on, umm, grammar and vocabulary"	"ah vocabulary……grammar"	"umm, I try to plan same as before"	Grammar and vocabulary
220	Which	3	"ahh, for example, the car which he, the car which he"	N/A	"same"	Grammar
221	Grammar	2	"ahh, last week I learn a grammar to you, today I also use grammar"	"ahh, last week I learn a grammar to you, today I also use grammar"	"umm, not different"	Grammar
222	Colors	3	"last week picture is, is, no colours but this picture has very colourful"	"ah story"	"ah, different"	Story
223	Vocabulary	3	"umm, I focus on vocabulary, umm, Vocabulary"	N/A	"same"	Vocabulary, grammar, story
224	Story	2	"I want to make, umm interesting story and I tried to remember, remember doing last week"	"story and grammar"	"not different"	Story and Grammar
225	Same	2	"umm, grammar is same"	"umm, story"	"ahh same"	Grammar and story
226	Is	4	"last time is easy"	"umm, story…..to use am or one day vocabulary"	"last time is easy"	Story and Vocabulary

At week three, as with the GP group, there were slightly less frequently used words from most of the GUP participants compared to week two. Again as with the GP group, question two was only used on a few students as the researcher relied more on question three *"what did you focus on when you were planning?"* which was similar in meaning and elicited greater responses from the participants, thus question two was not relied upon during most the interviews at weeks three and four. However, the omission of question two should have been cancelled out by the inclusion of question five *"did you plan differently compared to last week?"* which was used from week three to find out if the participants' strategies had changed from the previous week. Nevertheless, overall, there was still a slight drop in the most frequently used words from the GUP group.

At week three, the GUP group received unguided planning. In terms of their post-task responses, there was more varied selection of frequently used words from most of the participants when describing their planning strategies, however the most commonly used words amongst the GUP participants were *'grammar"* and *'which'*. After reviewing the example phrases provided by each participant, the common theme of the GUP group appeared to be *'learning grammar' or 'using same grammar''.* Once again this was supported by the results of the most frequently used words of the GUP group as a whole during week three's post-task interviews (see table 56). The word *'grammar'* was used the most by the GUP group at 14 occurrences.

Table 56. *Most frequently used words of the GUP group during week three*

Word	Frequency
Grammar	14
Is	11
Which	10
Story	10
To	9
Same	7

Eleven participants either mentioned a planning focus towards grammar during questions two or three, or mentioned grammar as an additional planning strategy when prompted in question four. One participant (222) attended towards the story only, and

one learner (226) focused on the story and vocabulary. However, nine learners referred to the storyline as an additional planning strategy. Finally, most of the participants except 215 and 222 confirmed that they used the same planning strategies in week three as they did in week two. Most of the GUP group commented on their continual attention towards the targeted forms, for example, *"umm, grammar is same"* and *"I use which and thinks".* Although 215 said that she planned differently compared to week two because in week three, planning was cognitively easier, *"this is easier…umm, I learn grammar".* The same applied to 226, *"last time is easy*". In addition, 222 said that his planning was different to week two on account that the narratives in week three involved colour, *"last week picture is, is, no colors but this picture has very colourful".*

We can therefore conclude that the GUP group's main planning strategy during week three was to use the targeted grammar forms, even though there was no grammar guidance provided. Furthermore, nine participants also attended to the storyline as an additional planning strategy whilst a smaller number attended to vocabulary. Finally, most of the GUP group confirmed that they used the same planning strategies as in week two.

7.5.2 Week four: GUP group's planning strategies

The planning strategies for the GUP group during week four are displayed in table 57. It displays the participant's identification number, the most frequently used word (N/A implies no word was frequently used), an example phrase, additional planning strategies (N/A implies there was no additional strategy). It also includes responses to question five which examined whether the learners' planning strategy had changed from the previous week. Finally, an interpretation of the learner's planning strategy is provided.

Table 57. *Week four: GUP group's planning strategies*

GUP group	Frequently used word	Frequency	Example	Additional planning strategy? (Question 4)	Different from last week? (Question 5)	Overall planning focus
214	Story	2	"story, story"	N/A	"almost"	Story
215	N/A	N/A	"story"	N/A	"erm no"	Story and grammar
216	Same	5	"same same"	N/A	"same"	Story and grammar
217	Or	3	"for example, green tyre or red white or open car or"	N/A	"type of writing is same"	Grammar and vocabulary
218	Which	2	"umm, which toka, she wants the car which have"	"umm vocabulary"	"umm same"	Grammar, story and vocabulary
219	Which	2	"umm, I used which and umm"	N/A	"same, almost"	Grammar and story
220	Grammar	2	"ahh, cars color, grammar"	N/A	"same"	Grammar
221	Same	2	"for example, erm erm erm for example sarah thinks she wants er it is same I use same grammar last week"	N/A	"erm not different"	Grammar
222	story	3	"last week, I focus on colour but today I focus on story"	N/A	"last week, I focus on color but today I focus on story"	Story
223	Same	2	"same"	N/A	"same"	Grammar, story, vocabulary
224	Grammar	2	"grammar"	"er er story and grammar"	"same"	Story and grammar
225	think, do	2	"er last time is more difficult I think"	N/A	"er last time is more difficult I think"	Vocabulary
226	Car	4	"there are four sisters er each sister thinks cars kind of car"	"story"	"same"	Story and vocabulary

At week four, as with the GP group, there was a slight increase in the most frequently used words compared to week three. Question four again was only used on a few students who did not provide a response to question three. If the students answered question three i.e. *'what did you focus on when you were planning?'* question four was considered unnecessary for the same reasons as with the GP group: the GUP learners could explain their planning strategies without the need for further prompting.

At week four, even though the GUP received unguided planning, the most commonly used words across the group were similar to weeks two and three i.e. *'which'* and *'grammar'* however the most frequently used word amongst the group was *'same'*. After reviewing the example phrases provided by each participant, the common theme of the GUP group again appeared to be use of the targeted grammar for example, *'same grammar' or 'using which'* etc. This theme was generally supported by the results of the most frequently used words of the GUP group as a whole during week four's post-task interviews (see table 58). The word *'same'* was used the most by the GUP group at 16 occurrences.

Table 58. *Most frequently used words of the GUP group during week four*

Word	Frequency
Same	16
And	14
Story	11
Car	10

Nine participants either mentioned that their planning focus was directed towards grammar during question three, or that their planning strategy was similar to the previous week which also focused on grammar. Two participants (214 and 222) focused on the story only during planning, one learner (225) focused on vocabulary, and one participant (226) focused on the story and vocabulary. However, as in week three, a large number of learners (eight) referred to the storyline as an additional planning strategy. Finally, most of the learners except for 222 and 225 confirmed that they used the same planning strategies that focused on grammar in week four as they did in week three, for example, *"erm for example sarah thinks she wants er it is same I use same grammar last week", "umm, to explain sisters I use the colors she there wearing and I*

used which and umm." Participant 222 however, said that he planned differently compared to week three because "*last week, last week, I focus on color but today I focus on story"* and when asked why, he responded, *"Ah, difficult, ah I, difficult to know this story"*. This suggests that the task at week four which had increased in complexity along intentional reasoning demands by having extra objects within the narrative (in this case cars) may have required extra effort to describe using the targeted RC types, thus forcing the learner to focus more on the story in order to describe the pictures. On the other hand, participant 225 commented that *"er last time is more difficult I think"* when asked why, she responded, *"because sometime I do this task maybe I think could do better"* which could imply that she was not as motivated to plan during week three compared to week four.

We can therefore conclude that the main planning strategy of the GUP group during week four was to once again use the targeted grammar forms even though they received unguided planning. Also, eight learners used the storyline as an additional planning strategy whilst a smaller number attended to vocabulary. Finally, most of the GUP group confirmed that they used the same planning strategies as in week three.

7.5.3 Conclusion of the GUP group's planning strategies

Despite the results of the GUP group's pre-study questionnaire in which half the group indicated a preference towards oral communication over accuracy, the above analysis showed that the planning strategies of the GUP group during weeks two, three and four appeared to remain largely focused towards grammar. Although there was a gradual decline in attention to form when they received unguided planning during weeks three and four of the treatment. The most frequently used words during the treatment indicated attention towards grammar, as did the common theme of the group's planning strategies, although during weeks three and four, most of the learners were using the storyline as an additional planning strategy. Finally, most of the GUP group confirmed that their planning strategies remained largely unchanged as the tasks increased in complexity. These results confirmed that the GUP group initially focused on grammar and then largely maintained their attention on form as they prepared for more complex narratives over time. In the following chapter, we discuss the findings of these results in detail.

8. DISCUSSION

8.1 Introduction

This chapter first discusses the results of each of the measures used in chapter six to assess the effects of guided planning and task complexity on L2 oral development. We then move onto discuss the results in chapter seven which relate to the planning strategies of the GP and GUP groups during their task sequencing treatments. This chapter will therefore begin by discussing the findings of each of the hypotheses that relate to research question one:

- To what extent does guided planning and task complexity facilitate L2 oral development in terms of fluency, morphological accuracy of OS and OPREP English relative clauses and 3rd person singular and plural, and syntactic complexity for second year Japanese university learners of English?

- Hypothesis one: guided planning and task complexity leads to less developmental gains in fluency compared to guided and unguided planning and task complexity.

- Hypothesis two: guided planning and task complexity leads to greater improvements in morphological accuracy of OS and OPREP English relative clauses and 3rd person singular and plural compared to guided and unguided planning and task complexity.

- Hypothesis three: guided planning and task complexity leads to greater gains in syntactic complexity compared to guided and unguided planning and task complexity.

- Hypothesis four: guided planning and task complexity leads to greater improvements in learners' receptive awareness of OS and OPREP English relative clauses and 3rd person singular and plural compared to guided and unguided planning and task complexity.

As research question one investigates L2 oral development, we begin in 8.2 by reviewing Kormos' model (2011) of L2 speech production from 2.3.1 which will help us to interpret the effects that the task sequencing treatment had on L2 oral development. From 8.3 to 8.6 we then discuss each of the above hypotheses in turn. Sub-section 8.7 then explains the findings of hypotheses one to four in relation to second language acquisition. In 8.8 we move onto discuss the findings concerning research question two:

- What strategies do Japanese second year university learners of English use when planning for oral narratives that increase in complexity over time?

8.2 Kormos' (2011) bilingual model of L2 speech production

As mentioned in 2.3.1, Kormos (2011) provides a bilingual model of L2 speech production which accounts for how L2 speech is produced, and it consists of four main components. The first stage is *conceptualization* which involves planning the goal of a communicative message, otherwise known as macro-planning and then deciding on the viewpoint to express it, known as micro-planning. At this stage, the message is referred to as a preverbal plan. The second stage of L2 speech production is *formulation* which concerns the grammatical and phonological encoding of the preverbal plan. The communicative specifications of the plan activate the required lexical items within a learner's mental lexicon. The lexicon consists of 'lexemes' which represent a learner's knowledge of L1 and L2 word forms and phonological information, and 'lemmas' which relate to their morphological and syntactic properties. Syntactic encoding begins with the activation of the appropriate lemma, followed by encoding of phrases and clauses. The third stage of L2 speech production is *articulation* which involves receiving and executing the intended message as spoken language. Finally, there is a *self-monitoring* component which checks for errors as speech is generated and processed.

We now turn to discuss the effects guided planning and task complexity had on these psycholinguistic processes in terms of developments in fluency, accuracy and complexity.

8.3 To what extent does guided planning and task complexity facilitate L2 oral development in terms of fluency?

According to hypothesis one, guided planning and task complexity facilitates L2 oral development to a lesser extent than guided and unguided planning and task complexity in terms of fluency. The fluency results analyzed in 6.2.1 showed that hypothesis one was not confirmed. Both the GP and the GUP groups produced significant gains in fluency from the pre-test to the immediate post-test as well as from the pre-test to the delayed post-test, thus showing that both respective treatments had positive consequences for fluency over time. These results were not surprising as we would expect most forms of treatment to have an impact on learning in some way. What is of pedagogic interest are the results *between* both groups as it was GP group who produced significantly greater gains from the pre-test to the immediate post-test (M = 28.87, SD = 14.70) compared to the GUP group (M = 16.45, SD = 14.39) using the pruned speech rate measure 'syllables per minute excluding repetitions, false starts, L1 use and incomprehsible language'. This difference was significant $t(22) = 2.09$, $p < 0.05$, and it represented a large effect (d = 0.89) between both groups. Similar results were also found from the pre-test to the delayed post-test as the GP group's mean gain (M = 28.47, SD = 13.36) was greater than the GUP group (M = 19.00, SD = 8.40). This difference was also significant $t(22) = 2.08$, $p < 0.05$ with a large effect size (d = 0.89). Consequently, guided planning and task complexity had a greater effect on fluency over time compared to guided and unguided planning and task complexity using the pruned speech rate b measure. These findings disproved our hypothesis which was based on the pilot study results in which the B2 guided planners produced *less* gains in fluency compared to the unguided planners. However, the discrepancy of the pilot and main study results for fluency probably came from the difference in group sample sizes as the pilot group sizes were so small (n = 2) that it may not have been possible to make precise hypotheses regarding a larger population.

The main study fluency findings did support the majority of previous strategic planning studies discussed in 3.3.5 which showed that guided or unguided planning leads to gains in fluency, with the exception of Mochizuki & Ortega (2008). Although, an important point to consider when comparing fluency findings with previous studies are the measures used to capture it. For example, as we saw in 2.4.2 there are temporal

speech rate measures for fluency, such as the number syllables per minute, as well as repair measures such as number of false starts or repetitions. As this study used a speech rate measure the findings are more comparable with studies that also used temporal measures. Those studies that did were Yuan & Ellis (2003); Sangarun (2005); Gilabert (2007a) who all used 'syllables per minute' and reported gains in fluency. These studies also used intermediate level learners, adding further similarities to the present study which used B2 intermediate level learners.

The unique aspect about the findings of this study is that it appears to be the only study that has investigated the effects of strategic planning over time. The only previous study that investigated task planning over time was Bygate (2001b) which involved repeating the same narrative task over a ten week period which led to significant gains in fluency using similar temporal measures of 'unfilled pauses per t-unit'. However, that study did not involve *strategic* planning where learners were allowed time to plan before performing a task. Rather, Bygate's (2001b) study involved repeating the same task type without the opportunity to plan prior to performance. Consequently, this study sheds new light on how task sequencing that involves two types of strategic planning: continuous guided planning, and guided planning followed by unguided planning can generate significant improvements in fluency over time with the former treatment having a greater effect.

Let us now examine why guided planning and task complexity resulted in greater gains in fluency compared to guided and unguided planning and task complexity. Clearly, gains in fluency from both groups occurred from having time to plan and rehearse the targeted forms as well as practice using them through tasks that increased in complexity. According to the claims of Robinson's (2003) Cognition Hypothesis in 3.4.3 increasing the intentional reasoning demands of tasks places greater linguistic demands on learners which 'pushes' their output. The tasks at weeks three and four of the treatment increased in complexity along intentional reasoning demands by containing additional obligatory cases of RCs and accompanying use of 3rd person singular or plural which is likely to have forced the learners to produce more instances of the targeted forms. For example, at week three, GUP participant 226 commented:

I: S*o when you were planning now, did you plan different to the last time?*
S 226: *Last time is easy*
I: *It's easier? Why?*

S 226: *Ahh, umm, to explain the people is, last time is little*
I: *A little?*
S 226: *Yes*
I: *Can you explain that bit more?*
S 226: *Ahh, like this is, many many focus, many fact*

Furthermore, at week four, GP participant 105 commented:

I: *Did you plan different compared to last week?*
S 105: *Erm Yes*
I: *Different OK and why*
S 105: *Why, I think this picture become more difficult*
I: *Ok for example?*
S 105: *For example Erm picture two cars er*

The effect of this task sequencing treatment is likely to have resulted in the learners memorizing the target forms in a way they might memorize formulaic language. In 2.3.2, Kormos (2011) reminded us that "the majority of our utterances are memorized phrases, clauses and sentences which together are called formulaic language" (p. 46). We saw that formulaic language can comprise of communicative functions such as apologizing and are generated in conceptualization as 'chunks' which contain multiple concepts that activate corresponding linguistic chunks within the lexicon that are stored as one lemma. Thus formulaic language is "produced faster and with less conscious effort than creatively-constructed elements of the message" (Kormos, 2011, p. 46). According to Kormos' (2011) model, practice opportunities facilitate the encoding of words and their associated syntactic information in the learner's mental lexicon and this assists the automatisation of the formulation process. In the case of this study, repeated attempts at producing RC types during the sequence narratives are likely to have strengthened the retrieval links of the required syntactic information which results in "more efficient message planning and faster lexical access and selection" (Gilabert, 2007b, p. 64). In this case, the continual process of planning for intentional reasoning speech during the treatment seems to have helped 'speed-up' lexical access and selection of the targeted forms and accelerated the production of the targeted language during the treatment narratives benefitting fluency. By the time both groups performed

the post-tests, it seems that they had automatised the L2 structures associated with intentional reasoning speech involving the targeted RC types and 3rd person singular or plural into formulaic chunks. For example *'which has', 'which have', 'looking at', 'he thinks he likes', 'she believes she likes'* etc. At the immediate and delayed post-tests, both groups were therefore able to conceptualize intentional reasoning 'chunks' as they explained the actions of other people which were then encoded by the corresponding lemmas. The proceduralisation of the targeted forms enabled both groups to produce the forms at a faster rate during the post-tests despite the lack of planning time available.

Furthermore, from weeks two to four the tasks increased in complexity along resource-dispersing dimensions as planning time was gradually reduced (from ten minutes at week two, to seven minutes at week three, to four minutes at week four) in order to increase "the ability to access and deploy knowledge during *performance* of a complex skill" (Robinson, 2005, p. 7). In other words, the reduction in planning time during the task sequence treatment would have primed learners to produce complex L2 speech under real-time conditions with no planning time. As the task sequence treatment progressed, both groups would have been required to access and produce L2 intentional reasoning speech at a faster rate due to the reduction in planning time. This treatment would have benefitted their fluency at the post-tests in which they had to produce complex language without planning time.

The significantly greater gains in fluency from the GP group appear to have resulted from the differences in the planning conditions of both groups at weeks three and four of the treatment in which the GUP group received unguided planning. This study argues that the improved performance in fluency from the GP group was attributed to the fact that the GP group received continual guidance towards the targeted forms which drew their attention to the production of the forms during the post-tests to a greater extent than the GUP group. For example, we know from the results in 7.3 and 7.5 that both groups' main planning strategy during week two was focusing on the targeted grammar forms, for example, *'using grammar' or 'using which have/has'*. However there were more instances of attention towards other areas of planning from the GUP group as the treatment progressed, for example, the storyline. Consequently, the GP group seemed to have had more practice opportunities at producing the targeted forms compared to the GUP group which may have aided the automatisation of the targeted forms to a greater extent than the GUP group, whom may have reverted to

using other linguistic forms to tell the story that were not automatised, and this resulted in a slower speech rate.

Ortega points out that attention to language form is "synonymous with a concern for being accurate" (2005, p.106) however, in this case, it is argued that attention to language form may have benefitted the GP group's fluency as well. As we know from Kormos (2011), formulaic language is produced at a faster rate compared to newly constructed messages. Thus, it is argued that the GP group were more aware to produce the targeted forms at the seven obligatory contexts of the immediate and delayed post-tests compared to the GUP group because the former group received continual grammar guidance towards them during the treatment and had more opportunities to process and automatise the language. As there were seven obligatory contexts in which to produce the targeted forms at the post-test narratives, this would have allowed the GP to rapidly produce formulaic language on several occasions which would have contributed to a more fluent performance compared to the GUP group who produced less instances of the targeted forms.

Having discussed the fluency results of this study, we now turn our attention towards the accuracy findings of the targeted forms.

8.4 To what extent does guided planning and task complexity facilitate L2 oral development in terms of morphological accuracy?

Hypothesis two claimed that guided planning and task complexity would produce greater gains in morphological accuracy of OS and OPREP RCs and 3rd person singular and plural compared to guided and unguided planning and task complexity. The accuracy results reported in 6.2.2 were confirmed. Both the GP and the GUP groups produced significant gains in terms of accurate use of the RC types OS and OPREP and 3rd person singular and plural from the pre-test to the immediate post-test as well as from the pre-test to the delayed post-test. This showed that both groups' task treatment had positive consequences for improvements in the accuracy of the targeted forms over time. In terms of differences between the groups, the GP group produced a greater gain from the pre-test to the immediate post-test, (M = 32.54, SD = 6.97) compared to the GUP group (M = 26.46, SD = 11.84) using the accuracy rating scale measure but it was

not significantly different $t(24) = 1.60$, $p > 0.05$ with only a medium effect ($d = 0.65$). However, from the pre-test to the delayed post-test, the GP group produced a greater mean gain ($M = 34.92$, $SD = 7.12$) compared to the GUP group ($M = 26$, $SD = 12.52$) and this difference was significant $t(24) = 2.23$, $p < 0.05$ with a large effect ($d = 0.91$). Consequently, guided planning and task complexity had a greater effect on the accuracy on the targeted forms over time compared to guided and unguided planning and task complexity using the rating scale measure.

The gains in accuracy of the GP and GUP groups appear to support the mixed results of the accuracy findings of previous planning studies that we discussed in 3.3.6. For example, the results of *guided* planning studies such as Sangarun (2005); Mochizuki & Ortega (2008) reported gains in accuracy, whereas the findings of *unguided* planning studies generally did not (for example, Yuan & Ellis, 2003; Gilabert, 2007a). In the present study, the GP group produced significantly greater gains in accuracy compared to the GUP group which lends weight to Mochizuki & Ortega's (2008) argument that "in order to maximise the effectiveness of planning time some sort of guidance in beneficial, particularly when increased accuracy is the goal" (p. 15). As the GUP group only received guided planning at the start of their treatment whilst the GP group received continuous guided planning throughout suggests that developments in accuracy correlate with the amount of guided planning provided. Furthermore, as Mochizuki & Ortega (2008) reported gains in accuracy of relativization from guided planning with beginner-level learners whilst the present study used B2 intermediate learners suggests that guided planning benefits accuracy with different proficiency levels. Both of these studies were also unique in terms of the gains afforded by guided planning on specific linguistic forms. Whilst Mochizuki & Ortega (2008) appeared to be the first study that showed the benefits of guided planning on RC production in natural language use, as opposed to using controlled tests, the findings of the present study were able to report the benefits of guided planning on RC production in natural language use *over time*.

Another important factor that contributed to the gains in accuracy of the present study was the specific measure used, in this case, the rating scale that was sensitive enough to track learners' use of the targeted RC types and 3rd person singular and plural. Specific measures are an important factor that can help us interpret the accuracy findings of this study in relation to the mixed results of previous studies. For example, previous planning studies (Yuan & Ellis, 2003; Bygate, 2001b) used general measures for accuracy such as 'errors per t-unit' which both Bygate (2001b) and Robinson (2007)

have argued may not be sensitive enough to track changes in accuracy and may have accounted for the null accuracy effects of those studies. Consequently, improvements in accuracy may depend not only on the type of strategic planning (guided or unguided) but also on the measures used.

As the pre-test narrative involved no planning time or guidance towards the targeted forms it was not surprising that the GP and GUP groups completed the tasks without using relativization and produced low means (*M* = 7.15, *SD* = 2.30) and (*M* = 6.85, *SD* = 2.54) respectively. However, as discussed in the previous sub-section, we know from the GP and GUP group's planning strategies in 7.3 and 7.5 that the guided planning conditions during week two of the treatment drew both groups' attention towards the targeted forms in order to explain the actions of other people using the RC types OS and OPREP as well as 3rd person singular and plural. In terms of Kormos' (2011) model, both groups at week two would have had time to rehearse using the forms and to check for errors through their monitoring system which we assume would have positive consequences for accuracy. Repeated practice opportunities at producing the targeted forms during the sequence narratives would have benefitted formulation by strengthening the retrieval links of the required syntactic information enabling them to produce the targeted forms at the post-tests. The automatisation of the required L2 structures associated with the targeted forms would have enabled the learners to access them as formulaic chunks during the seven obligatory contexts of the immediate and delayed post-tests benefiting accuracy. For example, as we saw in 6.2.2 the GP group's immediate post-test mean (*M* = 39.69) translated into 66.15% target-like accuracy in terms of the seven obligatory cases of RCs and the accompanying use of 3rd person singular or plural. The GUP group produced a smaller mean (*M* = 33.31) which converted into 55.52% target-like accuracy. At the delayed post-test, the quality of the targeted forms produced actually increased for the GP group (*M* = 42.08) which converted into 70.13% target-like accuracy. However, the GUP's accuracy decreased slightly (*M* = 32.85) with 54.75% target-like accuracy. Consequently, these results show the benefits of form-focused instruction in terms of accuracy development as the effects of guided planning and task complexity remained constant over time. According to Schmitt (2010) "a delayed post-test of three weeks should be indicative of learning which is stable and durable" (p.157). As this study incorporated a design in which the delayed post-test occurred three weeks after the last treatment session suggests that

guided planning and task complexity facilitated long-term learning effects in relation to the accuracy of the OS and OPREP RC types as well as 3rd person singular and plural.

The cause of the GP group's significant gains in accuracy appears to have been the guided planning conditions which, as in the case of the fluency findings, would have directed learners' attention towards practising the targeted forms to a greater extent than the GUP group whom may have used other linguistic forms to narrate the tasks. This was confirmed by the results in 6.2.2 which displayed the GP and GUP group's production of RCs at each of the seven obligatory contexts per narrative at the post-tests. As mentioned in 5.3.1, each narrative contained 5 OS RC types and 2 OPREP RC types. In terms of the immediate post-test, the GP group produced 73.63% of the expected OS and OPREP RCs. The GUP group did not produce as many of the expected RC types (58.24%). In terms of the delayed post-test, the GP group again produced 81.32% of the expected RC types. The GUP group however, produced only 59% of the expected RC types. Thus, the significant accuracy gains of the GP group were on account of more accurate production of the targeted RC types at the post-tests compared to the GUP group.

The accuracy gains of the GP group could lend weight to the findings of Erkman et al. (1988); Doughty (1991) outlined in 3.5.5 who argue that RC instruction towards more marked RC types such as OPREP can result in improved performance of less marked RC types such as OS. The GP group received more form-focused instruction towards the OPREP RC type during the course of their task sequencing treatment compared to the GUP group who only received guidance towards the OPREP RC at week two. If we refer to Ortega's (2005) argument outlined in the previous sub-section that a focus on form is synonymous with accuracy, then the extra grammar guidance that the GP group received during weeks three and four seemed to be the contributing factor for the additional gains in accuracy. The findings of this study showed that in order to *maximize* developmental gains in the accuracy of specific linguistic forms, continual guidance during planning may be necessary for intermediate level learners as they progress with oral tasks. Given Japanese university learners' declarative knowledge of RCs, it appears that practice opportunities to plan independently in order to produce the structures are not as effective for developing accuracy as continual guided planning. Having reviewed the accuracy findings of this study, we now discuss the complexity results.

8.5 To what extent does guided planning and task complexity facilitate L2 oral development in terms of syntactic complexity?

According to hypothesis three, guided planning and task complexity leads to greater gains in syntactic complexity compared to guided and unguided planning and task complexity. Hypothesis three was partially confirmed. The complexity results from 6.2.3 followed a similar pattern to the accuracy results of the previous section which is not surprising as both aspects of L2 speech involved using specific measures related to RCs. For example, at the pre-test narrative, only a small number of relative clauses per AS-unit were produced by the GP and GUP groups (M = 0.06, SD = 0.17) and (M = 0.04, SD = 0.10) respectively. This again was attributed to both groups' avoidance of the form. However, as discussed in the above fluency and accuracy sections, we know that both groups' attention was drawn towards the RC types at week two of the treatment and that both groups consistently planned to produce the RC types during weeks three and four of the treatment. As in the accuracy findings, repeated practice opportunities at producing RCs during both groups' task sequencing treatment would have enabled both sets of learners to retrieve and produce RCs as formulaic chunks at the immediate post-test with positive consequences for complexity. For example, the GP group produced a larger mean gain from the pre-test to the immediate post-test (M = 0.70, SD = 0.22) compared to the GUP group (M = 0.62, SD = 0.25) but this difference was not statistically significant $t(22) = .83$, $p > 0.05$. In addition, (d = 0.35) indicated a small to medium effect between both two groups. In terms of the pre-delayed post-test, GP group's gain remained the same (M = 0.70, SD = 0.19) whilst the GUP group's gain increased (M = 0.70, SD = 0.35) in line with the GP group, thus there were no significant differences $t(22) = -0.06$, $p > 0.05$, and there was no effect size (d = 0). As a result, guided planning and task complexity was only marginally more effective at producing complex output in terms of relative clauses per AS-unit compared to guided and unguided planning and task complexity.

These findings shed new light on how tasks can be sequenced with different types of strategic planning: continuous guided planning towards form, as well as guided and unguided planning to produce long-term gains in complex output concerning RCs. These findings also support the claims of Robinson's (2003) Cognition Hypothesis which states that tasks sequenced according to an increase in their cognitive demands

facilitates "optimal task-based L2 language use and learning opportunities over time" (Robinson, 2010, p. 242). It was evident that more practice using RCs by performing tasks that increased in complexity resulted in the automatisation of the target language as more RCs were produced at the post-tests in unplanned conditions. In addition, this study supports previous guided planning studies discussed in 3.3.5 which also reported gains in syntactic complexity, for example, Kawauchi, 2005; Sangarun, 2005; Mochizuki & Ortega (2008), the latter study using a similar complexity measure 'relative clauses per t-unit' to the present study. As this thesis was investigating the same RC type (OS) as Mochizuki & Ortega (2008), a similar specific measure was used in relation to the targeted form 'relative clauses per AS-unit', and as a result, it was able to track improvements in complex output that related to relativization. Consequently, the findings of this study suggest that strategic planning can have a more positive impact on complexity if learners' attention is directed towards a specific linguistic form *and* specific measures are used to track learners' production of it.

Unlike Mochizuki & Ortega (2008) which used beginner level high school learners, the participants in the present study were intermediate university learners of English who were expected to have more stable declarative knowledge of RCs. Consequently, these learners may only have needed initial guidance towards the targeted forms as the GUP group were able to utilize their existing knowledge to plan effectively and independently during weeks three and four of the treatment which enabled them to produce RCs during the post-test narrations. If this was the case, how then do we account for the significant differences between both groups' pre- delayed *accuracy* mean gains when both groups produced similar pre- delayed *complexity* mean gains involving RCs? This difference lies in the measures used. In the case of accuracy, we used a rating scale to measure learners' *accurate* production of RCs plus accompanying use of 3rd person singular or plural, whereas in the case of complexity, we were only interested in learners' output of RCs. Thus, in contrast to the accuracy findings, it appears that continuous guided planning towards RCs may *not* be necessary with intermediate level learners in terms of *complex* production. What does appear necessary however, are opportunities to practice using the form during narrative production which is where the benefits of Robinson's (2003) Cognition Hypothesis come into play. Sequencing tasks according to increasing intentional reasoning demands provide opportunities for learners to direct their attention and efforts at conceptualizing and producing more complex output in order to meet the demands of complex tasks.

Given the significant gains in both groups' L2 output, it appears that learners of intermediate proficiency are able to capitalize on the learning opportunities afforded by tasks that increase in intentional reasoning demands. The implications of these findings in relation to syllabus design and pedagogy within Japanese contexts will be picked up in the next chapter.

We now turn to discuss the results of the final hypothesis for research question one: namely learners' receptive awareness of the targeted forms.

8.6 To what extent does guided planning and task complexity facilitate L2 development in terms of learners' receptive awareness of the targeted forms?

Hypothesis four claimed that guided planning and task complexity leads to greater improvements in learners' receptive awareness of OS and OPREP RCs and 3rd person singular and plural compared to guided and unguided planning and task complexity. Hypothesis four was partially confirmed. The results of the grammatical judgement tests in 6.2.4 showed that both groups produced significant gains from the pre-tests to the post-tests as a result of their respective treatment. GUP group produced a larger percentage mean gain from the pre-test to the immediate post-test (M = 16.74%, SD = 15.74%) compared to the GP group (M = 16.04%, SD = 10.06%) but it was not statistically significant $t(24) = -.135$, $p > 0.05$ and resulted in a very small effect (d = 0.06). However, the GP group produced the greater percentage mean gain from the pre-test to the delayed post-test (M = 16.41%, SD = 10.79%) compared to the GUP group (M = 12.95%, SD = 12.02%) but again this difference was not statistically significant $t(24) = .772$, $p > 0.05$ whilst (d = 0.32) confirmed a small to medium effect between the groups. These findings imply that guided planning and task complexity, as well as guided and unguided planning and task complexity both produced significant and fairly similar transferable effects onto a different type of test regarding learners' receptive awareness of RCs and 3rd person singular and plural. In other words, both sets of treatment were geared towards L2 *oral* development in terms of fluency, accuracy and complexity yet the benefits afforded by both task sequencing conditions were also able to facilitate significant improvements in both groups' *receptive* awareness of the targeted forms.

One implication of these findings is that both groups scored quite highly on the pre-test (72.18% for the GP group, and 65.55% for the GUP group). This contrasts with the oral narrative accuracy pre-tests that we discussed in 8.4 in which both groups avoided the use of the targeted forms despite having seven obligatory cases to use them. Although the learners were free to use their full linguistic repertoire during the pre-test narrative whilst the grammatical judgement test specifically targeted the forms, an implication of these findings could be that both groups' receptive knowledge of the forms were higher than their productive knowledge, as was found in Izumi (2003) who used both productive and receptive tests to target the acquisition of similar RC types. We know from 3.2.5 that Japanese learners' previous English education has been devoted more towards writing and grammar as opposed to speaking which might have accounted for the discrepancy in the pre-test scores of the grammatical judgement pre-test which was in written format. Nevertheless, the results from the grammatical judgement tests appear to be unique as according to Ellis (2009a), no previous planning study has been able to show that the effects of task planning can be transferred to a different type of task. The present study however was able to report how different types of strategic planning and task complexity that were focused towards L2 oral development also produced developments in learners' receptive awareness of the targeted forms.

Thus far we have seen how guided planning and task complexity produced gains in fluency, accuracy, complexity, as well as gains in learners' receptive awareness of the targeted forms. Consequently, the following section discusses the implications of these findings in relation to second language acquisition.

8.7 The effects of guided planning and task complexity on second language acquisition

At the beginning of this study in 2.4.7, Housen et al. (2012) argued that developmental changes of a learner's internal L2 system could be acknowledged through improvements in fluency, accuracy and complexity as follows:

- Fluency: the proceduralisation of L2 knowledge which allows the learner to access L2 resources with reduced time delays benefitting performance.

- Complexity: developments in L2 knowledge through learning new grammatical structures.
- Accuracy: modification of a learner's internal L2 system to meet target-like performance and eliminate errors during production.

In order to ascertain whether guided planning and task complexity could promote developmental gains in CAF a pre- post-test design was carried out. As we have discussed in each of the sub-sections of this chapter, guided planning and task complexity, as well as guided and unguided planning and task complexity both produced significant gains in fluency, accuracy and complexity that remained over time. In terms of overall gains in CAF, the findings showed that the effects of guided planning and task complexity were more significant in terms of developments in fluency and accuracy compared to guided and unguided planning and task complexity. There were no significant differences between the two groups regarding complexity and learners' receptive awareness of the targeted forms.

The only previous planning study to carry out this type of design was Bygate (2001b). A limitation of this study was that it did not "provide data that can easily speak to the effects of task planning on the acquisition of specific linguistic features" (Ellis, 2005, p. 28). Bygate's (2001b) design was measured in terms of general linguistic changes of fluency, accuracy and complexity, and as a result, it could not account for the development of specific linguistic features. The measures used in the present study however, were able to investigate both general linguistic change as well as specific linguistic change. For example, general linguistic change can be accounted by the gains in fluency which used similar speech rate measures to Bygate (2001b): 'syllables per minute'. Specific linguistic change can also be acknowledged by the gains in accuracy which used the RC and 3rd person singular and plural rating scale measure, as well as the gains in complexity which used 'relative clauses per AS-unit'. Let us now discuss each of these three dimensions in relation to L2 acquisition.

8.7.1 Effects on fluency

In terms of fluency, the findings of this study could claim that guided planning and task complexity led to general linguistic change in terms of fluency, as the GP group produced significant gains in the pre- post-tests compared to the GUP group. Although as we saw in 3.5.1, researchers are divided over the role of fluency development and the proceduralisation of L2 knowledge. On one hand, researchers (Housen et al., 2012; Kormos, 2011; Robinson, 2010) argue that fluency development occurs through the automatisation of L2 rules, knowledge and formulaic chunks (as discussed in our fluency findings in 8.3). Ellis (2009a) however, argues that "the development of fluency and the acquisition of linguistic knowledge are arguably separate phenomena" (p. 504). Ellis referred to Schmidt's (1992) account that there is "little theoretical support from psychology for the common belief that the development of fluency in a second language is almost exclusively a matter of the increasingly skilful application of rules" (p. 377) (cited in Ellis, 2009a, p. 504). In other words, it is doubtful whether fluency development is dependent upon the ability to proceduralize grammatical rules. Instead, Ellis (2009a) claims that fluency development "depends on extending exemplar-based knowledge" (p. 504). Skehan (1998) distinguishes rule-based knowledge from exemplar-based knowledge. According to Skehan, the former assumes that language learning occurs through the processing of rules, for example "the sequence 'MV' must always be followed by a vowel. In the latter case, exemplars, learning is interpreted as the accumulation of chunks" (p. 53). Exemplar-based learning does not rely on language rules but instead consists of accumulating formulaic chunks by matching new input against what the learner already knows. Skehan mentions that rule-based systems are commonly associated with explicit learning which involves "selective attention (to rules) and conscious induction of abstract rules, with such rules having a potential influence upon performance" (p. 54). Ellis (1994) (cited in Skehan, 1998, p. 55) argues that explicit instruction is most warranted with complex L2 forms in order to make the function of the forms salient to the learner which will consequently help to facilitate future exemplars. In the case of the present study, explicit instruction was carried out regarding the difficult RC type OPREP, as well as correct use of verb tense with RCs that contain singular or plural head nouns which results in the use of either 3rd person singular or plural. According to Skehan (1998), "fluency is achieved either through use

of exemplars (memory based chunks) or through use of rule-base systems to generate future exemplars which can then operate autonomously" (p. 60). In the case of the present study, as the GP and GUP groups were provided with explicit instruction on the use of complex L2 forms, it is argued that fluency development was achieved through a rule-based system where learners were able to proceduralise their declarative knowledge of the targeted forms through the provision of guided planning and the performance of tasks that increased in complexity. The practice opportunities afforded by this process resulted in the automatisation of the targeted forms into formulaic chunks or as Skehan (1998) refers to as 'exemplars' which would have benefitted fluency at the post-tests (as discussed in 8.3). Thus, the present study argues that the fluency gains of the GP group developed through the proceduralisation of a rule-based system which in turn developed into formulaic language. Fluency development did not take place independently from the acquisition of linguistic knowledge as Ellis (2009a) claims, but rather, the GP group's internal L2 system was developed through the proceduralisation of their L2 knowledge, as outlined in Housen et al. (2012).

8.7.2 Effects on accuracy and complexity

Ellis (2009a) and Housen et al. (2012) are in agreement about improvements in accuracy and complexity as indicators of acquisition. As we saw in 3.5.1, Skehan (1998) claims that strategic planning can facilitate the restructuring of existing L2 knowledge due to the positive consequences it has on L2 complexity. This has led Ellis (2009a) to conclude that "more complex production will lead to acquisition" (p. 504). As the results of the present study produced significant gains in syntactic complexity for the GP and the GUP groups that did not diminish over time, it is argued that both groups' L2 knowledge may have been restructured as a consequence of their respective treatment and that acquisition may have occurred in terms of relativization using the 'relative clauses per AS-unit' measure. Although the complexity findings of this study are limited by the non-normal distribution of the data set reported in 6.2. Furthermore, as there were no significant differences between both groups neither treatment can claim to be more effective than the other.

The present study does claim that acquisition occurred with regards to the accuracy findings of the GP group. As discussed in 3.5.1, Ellis (2009a) refers this type

of acquisition as "the development of greater control (accuracy) over existing linguistic features" (p. 504). Both the GP and the GUP groups produced significant gains in the accuracy of the targeted RC types OS and OPREP as well as use of 3rd person singular and plural from the pre-test to the immediate and the delayed post-tests. However, the GP group produced significantly greater gains from the pre-test to the delayed post-test compared to the GUP group which showed that guided planning and task complexity benefitted learners' accuracy of the targeted forms to a greater extent over time, thus providing evidence of acquisition.

Although as we saw with the accuracy results in 6.2.2, both groups' low pre-test scores were not attributed to errors in the use of the targeted forms but rather learners' avoidance of the forms. The pre-test results support the findings of Schachter (1974), discussed in 3.5.4, which showed that Japanese learners may have a tendency to avoid using RCs due to difficulties in L2 production. Consequently, it could be argued that the accuracy gains produced from the pre- and post-tests were invalid because the GP and GUP groups were unaware to produce RCs at the pre-test, and as a result, the pre-test scores were not an accurate indication of the learners' ability to use RCs prior to their respective treatments. The issue of students' ability to perform tasks without using expected language is one of the weaknesses that tasks pose for L2 acquisition studies, and as a result, "many researchers have doubted the ability or desirability of using tasks to target particular features of language" (Samuda & Bygate, 2008, p. 122).

The present study however argues against these claims. Firstly, the purpose of this study was to investigate whether learners, after receiving guided planning and task complexity, would be able to produce complex L2 structures involving RC types and accompanying use of 3rd person singular and plural under real-world task conditions that involved no planning time. As mentioned in 3.5.2, the ability to produce grammatical structures during communication without thinking requires acquisition of implicit knowledge of the targeted grammar rules. "Implicit knowledge is intuitive, procedural, systematically variable, automatic, and thus available for use in fluent, unplanned language use" (Ellis, 2008, p. 418). Consequently, to confirm whether learners have acquired implicit grammar knowledge: in this case the targeted RC types, learners' free oral production would need to be assessed in a pre- post-test design in which the tests reflect conditions that do not allow the opportunity for conscious planning. Following previous studies that measured implicit knowledge (Ellis, 2005b) (cited in Ellis, 2008), this study used similar pre- post-tests which involved oral

narratives without planning time, and a timed grammatical judgement test. Both of these tests limited the opportunity for learners to engage in conscious planning. We can therefore assume these tests were reliable constructs that were specifically used to measure implicit knowledge of the targeted forms by creating conditions for unplanned language use, as recommended by Douglas (2001) in 5.3. As both groups were then able to produce significant gains in the accuracy of the targeted forms from the pre-tests to the post-tests, under these unplanned conditions, it is possible that acquisition of the targeted forms occurred. In other words, the GP group's explicit knowledge of the targeted forms had been explicitly learned and practised during the treatment which resulted in the automatisation of the forms which allowed the learners to produce them during the post-tests without conscious attention, thus confirming acquisition of implicit knowledge. The present study however, does not claim that the GP group acquired implicit knowledge of the targeted forms, but rather, as Dekeyser (2003) and Ellis (2008) point out in 3.5.2, the learners' explicit knowledge of the targeted forms had been proceduralized to the extent that they could access and produce the forms accurately under unplanned conditions. In other words, they had acquired automatized explicit knowledge, as given the learners' intermediate L2 proficiency there was probably some small degree of conscious attention when producing the targeted forms.

Furthermore, although the pre-test scores were disappointing in terms of learners' production of the targeted forms, we know from 3.5.4 that RCs are instructed to Japanese learners during the second year of junior high school. The participants were therefore aware of the forms but choose not produce them at the pre-test. In an attempt to counter the oral narrative's weakness in targeting learners' use of the RC types and 3rd person singular and plural at the pre-test, the grammatical judgement test was used as a controlled measure to target learners' receptive awareness of the targeted forms. This test was able to measure learners' accuracy of the forms prior to the treatment as well as the significant improvements of the GP group at the immediate and delayed post-tests. These results therefore support our claim that guided planning and task complexity resulted in the acquisition of linguistic knowledge in terms of oral accuracy of targeted L2 forms, as well as receptive awareness of the targeted forms.

Finally, another indication that acquisition occurred with the GP group lies in the results of the delayed post-tests. As discussed in 8.3, Schmitt (2010) argues that acquisition studies need to show that long-term language learning effects have taken place, and this can be verified by the results of a three week delayed post-test. The fact

that the GP group's results did not decline in performance from the immediate to the delayed post-tests in terms of fluency, accuracy, complexity as well as their receptive awareness of the targeted forms indicates that guided planning and task complexity resulted in long-term learning effects.

The pedagogic implications of these findings will be discussed in the next chapter. We now move onto the final section of this chapter which discusses the results related to research question two.

8.8 What strategies do Japanese second year university learners of English use when planning for oral narratives that increase in complexity over time?

Let us first discuss the findings of the GP group's planning strategies followed by the findings of the GUP group, and then finally comparing both groups' strategies over time.

8.8.1 The GP group's planning strategies

The main planning strategy of the GP group at week two of the treatment was attention towards grammar involving the targeted RC types and 3rd person singular and plural, and this planning strategy remained largely unchanged as they prepared for more cognitively demanding tasks over time.

At week two, the results from the post-task interviews in 7.3 showed that the most frequently used word from the GP group was *'grammar'* which occurred 26 times and the common theme of the group's planning strategies appeared to be *'using grammar'.* At week three, the tasks increased in complexity along intentional reasoning demands by containing additional contexts of the targeted forms within the storyline. The tasks also increased in complexity along resource-dispersing dimensions as planning time was reduced from ten minutes to seven minutes. However, the GP group still used the words *'which'* and *'grammar'* the most during their post-task interviews, at 16 and 13 occurrences respectively, whilst the common planning theme again was *'using grammar'* or *'using which'*. Finally, at week four, as the tasks further increased in complexity along intentional reasoning demands and also reduced in planning time from seven minutes to four minutes, the GP group's planning strategies appeared

unchanged as the most frequently used words were *'which'* on 14 occurrences and *'grammar'* on 12 occurrences whilst the common theme was again *'using grammar'* or *'using which have/has'*. Furthermore, the majority of the GP group also confirmed that their planning strategies remained largely unchanged as the tasks increased in complexity, thus verifying the GP group's serial attention to form over time.

The only previous study that has appeared to report learners' planning strategies using post-task interviews was Ortega (2005) however that study only examined learners' perceptions at a specific point in time. This study interviewed learners over a three week period and as a result was able to report the patterns that emerged with learners' planning strategies as they prepared for tasks week by week. Ortega's (2005) study also did not involve guided planning and so it was unable to comment whether learners respond to planning instructions as expected. The present study's findings however tell us that this sample of Japanese university learners responded to the guided planning instructions as expected. In the case of the GP learners, explicit instruction towards RC types and 3rd person singular and plural resulted in explicit learning of the targeted forms, and their attention towards the forms remained largely unchanged over time.

Once again referring to Ortega's (2005) argument that "if we take attention to form as being synonymous with a concern for being accurate and/or being sophisticated while using the L2" (p. 106) then the interview responses of the GP group provide strong evidence to suggest that these learners were focusing on form in order to speak accurately or to use more complex language during their task performances. As their planning strategies remained unchanged throughout the treatment, this would lead us to suggest that the GP group were predominantly using planning time to focus on the grammar guidance provided in order to improve their accuracy and complexity. Consequently, and as we would expect, these results show that when learners are provided with guidance towards grammatical structures that are intended to help them complete a task, their attention will be drawn towards it during planning and performance. Furthermore, learners will continue to serially attend towards the targeted forms as they plan for more cognitively challenging tasks over time.

The GP group's attention towards form over meaning would suggest that this sample of learners prioritized accuracy and complexity over fluency when planning. This could be due to the grammar guidance provided or that this sample of learners were more orientated towards speaking accurately as opposed to speaking fluently. The

results from the GP group's questionnaire in 5.6.3 which investigated their orientations towards L2 speech however, suggested that their attention towards form during planning was probably due to the grammar guidance provided. For example, the questionnaire showed that 54% of the group rated a preference towards communication over accuracy when speaking in the L2 whilst 46% of the GP group indicated a preference towards communication *and* accuracy when speaking in the L2. None of the participants indicated they were more orientated towards accuracy over communication. Consequently, these results tell us that individual learner differences did not seem to be as significant a factor in determining learners' planning strategies when under the influence of guided planning. In other words, as this sample of learners were provided with explicit grammar guidance intended to help them complete a task, they generally devoted their main planning strategy towards using the targeted grammar form regardless of their personal orientation towards communication.

Finally, although the GP group's main planning strategy was a focus on form, certain learners also attended to meaning as well by focusing on the communicative aspect of the task i.e. the storyline. For example, at week two, certain learners confirmed their additional planning strategy was the storyline in 7.3 whilst the word *'story'* was the second most frequently used word during the interviews on 12 occurrences. Consequently, these findings support Ortega's (2005) claim that more advanced learners attend to form-in-meaning strategies, that is, certain learners "seemed to pay attention to the inextricable relationship between form and meaning, simultaneously holding in long-term memory considerations regarding the message to be conveyed and the essential formal resources to convey it" (p. 106). In other words, strategic planning afforded time for learners to weigh up the communicative task demands *and* to attend to the language required to complete it. In Ortega's (2005) study, learners were aware of the communicative nature of the task (a story-telling narrative) yet they also focused on form during strategic planning without receiving any instructions to do so. Ortega's (2005) metaphor of attention to form-in-meaning is supported by DeKeyser, Salaberry, Robinson, and Harrington (2002) who also argue that "simultaneous attention to form and content is clearly possible" (p. 809). In the present study, particularly with the GUP learners in the following sub-section, there were numerous examples that support Ortega's (2005) metaphor of attention to form-in-meaning as certain learners commented that they attended to the targeted grammar forms whilst also using additional planning strategies that focused on the storyline in

order to prepare for the communicative demands of the task, in this case narrating a story. However, an important factor in Ortega's (2005) study and the present study's findings regarding learners' attention to form-in-meaning strategies when planning is the proficiency level of the participants. In both studies learners were of upper-intermediate B2 oral level, and as a result, had sufficient L2 explicit knowledge to be able to integrate form and meaning simultaneously. This might not be possible with lower level proficiency learners who might need to attend to form and meaning separately during planning as the cognitive demands of the tasks would probably be too high.

8.8.2 The GUP group's planning strategies

The main planning strategy of the GUP group was a focus on the targeted grammar forms which continued as the learners prepared for more cognitively challenging tasks even though they received unguided planning during weeks three and four. However, there was a gradual decline in attention to form as the treatment progressed where learners attended to additional aspects of planning such as the storyline. Initially, at week two, the most frequently used word from the GUP group's post-task interview results in 7.5 was *'grammar'* at 33 occurrences and the common theme of the group's planning strategy was *'using grammar'.* At week three, as the tasks increased in complexity, the GUP group still used the word *'grammar'* the most during their interviews at 14 occurrences whilst the common theme was once again *'using grammar'* or *'using which'*, although at this stage of the treatment, most of the learners were using the storyline as an additional planning strategy. At week four, the most frequently used word was *'same'* at 16 occurrences as the common theme of the planning strategies was *'using grammar'*. Finally, the post-task interviews revealed that most of the learners attended to the storyline as well. On the whole however, the GUP group confirmed that their planning strategies remained largely unchanged as the tasks increased in complexity over time so we can conclude that their main planning strategy throughout the treatment was a focus on the targeted grammar forms.

The GUP group provided more examples of attending to additional aspects of planning, other than grammar, compared to the GP group. This appeared to be due to their *unguided* planning conditions during weeks three and four in which the GUP

group could plan independently. During week three, although their main planning strategy was grammar, nine of the thirteen learners confirmed that an additional planning strategy was a focus on the storyline. For example, the word *'story'* was used 10 times compared to the most frequently used word *'grammar'* at 14 occurrences. Examples of GUP learners attending to the storyline include participant 224 *"I want make, I want to make, umm, interesting story"* whilst learners 218, 224 and 226 each provided an example of attending to the storyline whilst using the targeted grammar *"for example, simon wants the car which have red tyre", "his brother thinks he wants the car with…", "ahh, he wants to the car which has black window and tyre"*. This pattern of attending to the storyline as well as the targeted grammar continued into week four, where the word *'story'* was used 11 times in the GUP group's interviews compared to the most frequently used word *'same'* on 16 occurrences, whilst *'story'* was also used more frequently than the word *'grammar'* (7 occurrences). As in week two, the GUP learners gave examples of attending to the storyline whilst using the targeted forms, for example, participant 226 commented *"er four sisters er there are four sisters er each sister thinks cars kind of cars"* whilst 218 commented, *"she wants the car which have"*.

The integration of different aspects of planning as the GUP group's treatment progressed, in this case, grammar and the storyline, provides evidence of the GUP group attending to form-in-meaning where the learners were using both form and meaning planning strategies simultaneously. This planning strategy appeared more in line with the GUP group's oral orientation towards speaking in the L2. The results of their pre-study questionnaire in 5.6.3 showed that 54% of the GUP group indicated a preference towards communication *and* accuracy whilst 46% of the group rated a preference towards communication compared to accuracy. None of the participants indicated they were more orientated towards accuracy over communication. Although the learners clearly showed a preference towards communication and accuracy when speaking in the L2, this only became evident in their planning strategies after they had received unguided planning in weeks three and four. Consequently, these results show how the planning conditions of the GUP group influenced their planning strategies. When provided with initial guided planning at week two, the GUP learners obeyed instruction and focused on the targeted grammar forms. However, when provided with unguided planning at weeks three and four, the GUP group still maintained their attention towards form but there were more examples of attention towards form-in-

meaning as learners began focusing on the storyline as well as grammar which reflected their oral orientation towards communication and accuracy. The results of the GUP group's planning strategies suggest that when learners are provided with initial guidance towards grammatical structures that are intended to help them complete a task, their attention will be drawn towards it during planning. However when given the opportunity to plan independently for more cognitively demanding tasks over time, the majority of learners continue to serially plan towards the target grammar but adopt more form-in-meaning strategies which reflect their L2 oral orientation towards speaking fluently and accurately.

Having reviewed the findings of the GP and the GUP group's planning strategies, the final sub-section of this chapter compares both groups' strategies for similarities and differences.

8.8.3 Comparing the GP and GUP group's planning strategies over time

There were similarities between the GP and the GUP group's planning strategies. For example, in line with the findings in Ortega (2005), both groups appeared to use strategic planning "to utilize various funds of explicit knowledge that guided their conscious attention towards areas in which they were well aware of holes and gaps vis-vis the specific task demands" (p. 106). In this case, both groups appeared to acknowledge the value of the grammar guidance provided in helping them meet the demands of a narrative that required its use, and therefore consciously attended towards practicing the forms throughout the treatment.

The results of this study showed that both groups of learners were willing to follow the guided planning instructions and primarily focus on form, although this focus on form did not always match their personal communicative orientation towards L2 speaking. These findings appear to counter Ortega's earlier claim that:

> attention to form cannot be assumed as a guaranteed byproduct of pre-task planning opportunity and that the communicative requirements of the task at hand and learners' general predisposition toward communication or proficiency in the L2 can substantially affect decisions regarding conscious allocation of attention and effort. (1999, p. 136)

In the present study however, it is argued that B2 oral level learners' attention during strategic planning is ultimately dependent upon the planning conditions imposed upon them, in other words, whether strategic planning is guided or unguided. In the case of the former, there is more certainty that learners *will* attend to form. This is echoed by Ellis (2009a) who claims that "learners make up their own minds about how best to plan and will not be unduly influenced by the task-designer's instructions unless there are very specific (as in Mochizuki and Ortega 2008)" (p. 500). The present study followed Mochizuki & Ortega's (2008) planning conditions which were targeted towards specific RC types. As a result, both the GP and the GUP groups appeared to acknowledge the value of the grammar guidance provided at week two as they explicitly attended to the forms throughout the treatment regardless of their personal orientation towards L2 communication. However, B2 oral Japanese learners may also adopt form-in meaning strategies which can reflect their L2 oral orientation towards communication and accuracy if they are allowed to plan independently on subsequent task versions. The pedagogic implications of these findings will be discussed in the next chapter.

9. CONCLUSIONS AND IMPLICATIONS

9.1 Introduction

This final chapter is divided into four sections. We begin in 9.2 by reviewing the main findings that relate to research question one whilst 9.3 reviews the findings for research question two. In 9.4 we discuss the pedagogic implications of this thesis and 9.5 outlines the limitations of the study. Finally, 9.6 describes areas for future research.

This thesis consisted of two research questions, the first was:

- To what extent does guided planning and task complexity facilitate L2 oral development in terms of fluency, morphological accuracy of OS and OPREP English relative clauses and 3rd person singular and plural, and syntactic complexity for second year Japanese university learners of English?

Four hypotheses were devised and we shall now review the key findings of each of them.

9.2 Main findings of hypothesis one

Hypothesis one claimed that guided planning and task complexity leads to less developmental gains in fluency compared to guided and unguided planning and task complexity. Hypothesis one was not confirmed. The GP group produced significantly greater gains from the pre-test to the immediate post-test (M = 28.87, SD = 14.70) compared to the GUP group (M = 16.45, SD = 14.39) using the pruned speech rate measure 'syllables per minute excluding repetitions, false starts, L1 use and incomprehsible language'. This difference was significant $t(22) = 2.09$, $p < 0.05$, and it represented a large effect (d = 0.89) between both groups. Similar results were also found from the pre-test to the delayed post-test as the GP group's mean gain (M = 28.47, SD = 13.36) was greater than the GUP group (M = 19.00, SD = 8.40). This difference was also significant $t(22) = 2.08$, $p < 0.05$ with a large effect size (d = 0.89). Consequently, guided planning and task complexity had a greater effect on fluency over

time compared to guided and unguided planning and task complexity using the pruned fluency speech rate measure.

9.2.1 Main findings of hypothesis two

Hypothesis two claimed that guided planning and task complexity leads to greater developmental gains in morphological accuracy of OS and OPREP RCs and 3rd person singular and plural compared to guided and unguided planning and task complexity. Hypothesis two was confirmed. The GP group produced a greater gain from the pre-test to the immediate post-test, (M = 32.54, SD = 6.97) compared to the GUP group (M = 26.46, SD = 11.84) using the accuracy rating scale measure but it was not significantly different $t(24) = 1.60$, $p > 0.05$ and only a medium effect was reported between the groups (d = 0.65). However, from the pre-test to the delayed post-test, the GP group produced a greater mean gain (M = 34.92, SD = 7.12) compared to the GUP group (M = 26, SD = 12.52) and this difference was significant $t(24) = 2.23$, $p < 0.05$, whilst (d = 0.91) suggested a large effect. This showed that guided planning and task complexity had a greater effect on morphological accuracy of the targeted forms over time compared to guided and unguided planning and task complexity using the accuracy rating scale measure.

9.2.2 Main findings of hypothesis three

According to hypothesis three, guided planning and task complexity leads to greater gains in syntactic complexity compared to guided and unguided planning and task complexity. Hypothesis three was partially confirmed. The GP group produced a larger mean gain from the pre-test to the immediate post-test (M = 0.70, SD = 0.22) compared to the GUP group (M = 0.62, SD = 0.25) using 'relative clauses per AS-unit' but this difference was not statistically significant $t(22) = .83$, $p > 0.05$. In addition, (d = 0.35) indicated a small to medium effect between both two groups. In terms of the pre-delayed post-test, GP group's gain remained the same (M = 0.70, SD = 0.19) whilst the GUP group's gain increased (M = 0.70, SD = 0.35) in line with the GP group, thus there were no significant differences between the groups $t(22) = -0.06$, $p > 0.05$, and there was no effect size (d = 0). Consequently, guided planning and task complexity was only

marginally more effective at producing complex output in terms of relative clauses per AS-unit compared to guided and unguided planning and task complexity.

9.2.3 Main findings of hypothesis four

Hypothesis four claimed that guided planning and task complexity leads to greater developmental gains in learners' receptive awareness of OS and OPREP RC types and 3rd person singular and plural compared to guided and unguided planning and task complexity. Hypothesis four was partially confirmed. Although the GUP group produced a larger percentage mean gain from the pre-test to the immediate post-test (*M* = 16.74%, *SD* = 15.74%) compared to the GP group (*M* = 16.04%, *SD* = 10.06%) it was not statistically significant $t(24) = -.135$, $p > 0.05$ and resulted in a very small effect ($d = 0.06$) between the groups. However, the GP group produced the greater percentage mean gain from the pre-test to the delayed post-test (*M* = 16.41%, *SD* = 10.79%) compared to the GUP group (*M* = 12.95%, *SD* = 12.02%) but again this difference was not statistically significant $t(24) = .772$, $p > 0.05$ whilst ($d = 0.32$) confirmed a small to medium effect. These findings showed that guided planning and task complexity produced fairly similar transferable effects onto a different type of test that measured learners' receptive awareness of the targeted forms compared to guided and unguided planning and task complexity.

9.3 Review of the main findings: research question two

The second research question of this thesis was:

- What strategies do Japanese second year university learners of English use when planning for oral narratives that increase in complexity over time?

The results of the post-task interviews revealed that the main planning strategy of the GP group at week two was attention towards the targeted OS and OPREP RC types and 3rd person singular and plural, and this planning strategy remained largely unchanged as they prepared for more cognitively demanding tasks during weeks three

and four. Furthermore, the majority of the GP group also confirmed that their planning strategies remained largely unchanged as the tasks increased in complexity over time.

In terms of the GUP group, their main planning strategy was a focus on the targeted grammar forms which continued as the learners prepared for more cognitively challenging tasks during weeks three and four. The GUP group also used additional planning strategies to focus on the storyline, particularly when they were allowed to plan independently during weeks three and four. On the whole however, the GUP group confirmed that their planning strategies remained largely unchanged as the tasks increased in complexity over time.

9.4 Pedagogic implications of guided planning and task complexity vs guided and unguided planning and task complexity on L2 oral development

There are a number of pedagogic implications regarding the findings of this study. First, in terms of methodology, explicit instruction towards grammatical features followed by the performance of tasks relates to the type of methodology that we discussed in 3.2.4, known as 'task-supported language teaching' (TSLT). In this case, tasks support the pre-teaching of grammatical items and serve as a means to allow learners to engage in communication whilst using the forms. Although this methodology differs from the TBLT methodology that we saw in Willis' (1996) framework in 3.2.3 which advocates attention to form *after* task performance, it appears that TSLT may be a suitable option within a Japanese educational context where many learners are more accustomed to traditional methods of language instruction that focus predominantly on grammar translation. As mentioned in 3.2.5, many Japanese learners have had little practice using their spoken English in communicative situations during their English education. As a result, asking Japanese learners to engage in oral communicative tasks with no guidance towards the language required to help complete them may result in impoverished language use which Seedhouse (1999) claims TBLT is guilty of. The advantage of TSLT however, is that it can provide scaffolding for learners by providing the target language required to complete tasks. Learners can then rely on guided planning to assist them in performing the task, and as they continue performing tasks that increase in complexity, their guidance towards form can gradually be reduced as their knowledge

of the forms become proceduralised to point where they can perform the task under real-word conditions of unplanned language use. TSLT also has the added advantage of drawing learners' attention towards linguistic forms known for their difficulty in oral production and encouraging its use during task performance, as in the clear benefits of the present study with Japanese learners' use of RCs. It is therefore argued that TSLT would be a preferable option than TBLT with Japanese learners who have lacked exposure with oral English communication. However, as learners become more proficient users of the L2 they could then be exposed to TBLT once they have the explicit knowledge and confidence to perform unfocused tasks. Thus TSLT could serve as a pre-cursor to TBLT within Asian contexts.

Another implication of this study relates to the role of the teacher during the treatment. As the participants of this study were B2 intermediate level learners of English, they had already acquired declarative knowledge of RCs from their previous English language education as discussed in 3.5.3. Consequently, both groups only required planning time and practice opportunities to perform increasingly complex tasks in which to proceduralise the target language into automatised explicit knowledge. In order to draw learners' attention towards the targeted OS and OPREP RC types and 3rd person singular or plural, only a ten minute instruction period was required at the start of the treatment, after which, both groups were able to plan for oral tasks independently and perform the sequence narratives. No teacher assistance or corrective feedback was provided during the treatment. The results of the study showed that Japanese learners of intermediate proficiency are able to capitalize on the learning opportunities afforded by tasks that increase in complexity. The monologic nature of the narratives and the guided planning conditions allow the learners to plan and perform the tasks autonomously. Thus the role of the teacher during this task sequencing process would change from initial teacher-led instruction at the start of the treatment in order to draw learners' attention towards the targeted forms and faceplate the noticing of input. After which, the teacher would then switch to 'facilitator' as learners plan and perform the tasks independently, offering feedback or assistance where necessary. However, in order for Japanese learners to maintain their focus on form, tasks used would need to be focused in order to help elicit learners' use of the target language.

We now turn to the pedagogic implications of this study in terms of syllabus design. As TSLT is ultimately determined by linguistic forms, in the case of the present study it was OS and OPREP RC types and 3rd person singular and plural, task

sequencing of this kind appears suited towards a structural syllabus that consists of a list of grammatical features to be instructed. As a result, focused tasks can be designed and sequenced to support the delivery of specific forms via explicit instruction, and the linguistic forms comprise of the syllabus. As a result, the findings of this study appear promising in relation to how tasks can be successfully implemented within a Japanese educational context. For example, in 3.2.5 we discussed how MEXT has been concerned with traditional methods of English language instruction that focused on grammar translation. This was partly attributable to high school English language courses that used structural syllabi devoted towards atomistic grammar, reading and writing that was not seen as conducive for developing Japanese learners' English oral skills (Brown & Kikuchi, 2009). This appears to not bode well for TBLT as Ellis (2009b) notes:

> Educational systems in many parts of the world place the emphasis on knowledge learning rather than skill development, and a task-based approach to language teaching is not readily compatible with such a philosophy. A structural approach based on teaching discrete items of language accords more closely with such an educational philosophy. (p. 242)

In the case of Japan's educational context which is rooted in traditions of English language instruction through discrete linguistic items (Sakui, 2004), TSLT could provide an outlet for the use and implementation of oral tasks that could accommodate teaching practices that rely heavily on the use of structural syllabi. For example, linguistic features identified within a structural syllabus, such as RCs, could be practiced through guided planning and task complexity to develop learners' L2 speech. As we have seen in this study, RCs in particular can benefit fluency due to its clausal nature that combines word forms together, for example, *'looking at'*, *'which have'* as well as orientating learners' attention towards complimentary structures associated with them such as 3rd person singular, *'he likes the dog which has…'* Thus guiding learners' attention towards the form and having them practice the forms with focused tasks that increase in complexity helps to formulate formulaic chunks and improve fluency. In order for this to occur with other linguistic forms, focused tasks would need to be designed that could elicit their use, and as we discussed in 3.2.1, this is not an easy feat to achieve.

In terms of oral development, careful consideration would also need to be given to match linguistic features with learners' proficiency. Ellis (2003) notes that

"presenting and practising features learners have failed to use correctly in production may not result in their acquisition if the learners are not developmentally ready to acquire them" (p. 30). As we saw in 3.3.6, the results in Mochizuki & Ortega (2008) were disappointing in terms of the amount of RCs produced by the sample as a whole partly because the learners' beginner level proficiency was too low to capitalize on the planning opportunities provided. More positive results were reported in the present study which used B2 intermediate level learners who had the explicit knowledge to capitalize on the guided planning conditions and perform increasingly complex tasks that proceduralised their knowledge of the target language. Thus, an important factor for the success of TSLT would be for curriculum designers to match linguistic features with learners' proficiency in order to ensure optimal learning conditions.

Finally, we consider the implications of this study on a wider population of learners. As discussed in 3.4.2, the task sequencing treatment of this study is based on Robinson's (2003) Cognition Hypothesis which states that optimal gains in L2 development occurs through sequencing tasks from simple to complex. As the participants of this study were second year B2 intermediate Japanese university learners of English, we can therefore suggest that the findings of this study may be generalised to similar populations of learners. However, Ellis (2003) notes that task-based studies that are based on psycholinguistic models such as Robinson's (2003) Cognition Hypothesis have been challenged by another theoretical account of task-based language learning referred to as socio-cultural theory. This states that the effects of tasks cannot be generalised because learners perform tasks in unique ways according to their own motives and perceptions. For example, Coughlan & Duff (1994) compared the task performance of five learners (four Hungarian and one Cambodian) who performed a picture description task one-on-one with a researcher in different contexts. The Cambodian learner completed the task at the researcher's home during a one hour meeting whilst the Hungarian learners performed the task at a school during a shorter meeting that lasted 20 minutes. The task was intended to be a monologue yet the Cambodian's performance was more dialogic as he repeatedly interacted with the researcher by asking for clarification and assistance. In terms of the Hungarian learners, their performance was monologic as the researcher was required to complete the tasks within a certain timeframe and so did not engage in interaction. This led Coughlan & Duff (1994) to acknowledge the influential role which the researcher has in shaping learners' discourse. As the Hungarian learners were instructed to perform the task in a

monologue, their performances differed according to their interpretation of the task. For example, one learner described the contents of the picture whilst another learner compared the picture with her personal experiences. Different interpretations towards the same task resulted in different types of discourse from each learner. Furthermore, learners' interpretation of the task changed again when they repeated the task. The divergence in the learners' performance led Coughlan & Duff (1994) to question whether the picture description task represented a natural, real-world communicative activity, a criticism which Ellis (2003) claims a lot of task-based studies may also be guilty of. As a result, learners may perform tasks in different ways depending on their interpretations or motives. For example, some learners may enjoy communicating in a task whilst others may simply view tasks as speaking exercises. In conclusion, Couglan & Duff (1994) note that "while the task or blueprint may be the same, the activity it generates will be unique" (p. 190). In other words, tasks should not be considered constant, and may not have predictable effects for learning as learners react to them in different ways.

Differences in task performance as a result of learners' motives or interpretations regarding task goals is referred to as "task-as-workplan and task-in-process" (Samuda & Bygate, 2008, p. 49). The former represents a teacher's or researcher's intention of a task whilst the latter reflects the actual language produced by the learner. Seedhouse (2005) argues that language intended to be produced from a task (task-as-workplan) may not occur due to learners interpreting tasks in their own way which can result in unexpected L2 output (task-as-process). Seedhouse (2005) warns that "this can cause serious problems with validity in task-based research" (p. 176) as differences between intended language production and learners' actual task performance can be so great that it jeopardizes task-based research and pedagogy due to the uncertainty surrounding learners' L2 output.

In terms of the present study, the results suggest that guided planning and task complexity can produce language as intended by task designers. For example, as described in 5.3.1, each narrative contained 7 obligatory contexts of RCs consisting of 5 OS and 2 OPREP RC types. In 6.2.2, the results of the immediate post-test showed that the GP group produced 73.63% of the expected OS and OPREP RC types whilst at the delayed post-test, the GP group produced 81.32% of the expected RC types. Thus, these results show that when learners are provided pre-task guidance towards language required for a focused task, in this case relative clauses, learners generally do produce

language as intended by the researcher. As a result, this study argues that TSLT provides a way around the problem of task-as-workplan vs task-as-process. Specifically, using focused tasks designed to elicit specific language forms combined with pre-task guidance towards the target form can increase the probability of learners producing intended language features.

Another implication of the present study concerns the role of the researcher as a participant during task performance. As discussed in 5.3.1, each participant performed a narrative one-on-one with the researcher who acted as the listener. As the tasks were performed in a spare classroom in an experimental setting, it created a unique environment for the participants to speak in the L2. Thus, in line with Coughlan & Duff (1994), the task would probably not constitute a natural communicative activity as it is doubtful that the participants had experienced anything like this before. Consequently, if the same learners were to perform the tasks again with other students in a regular lesson, the different conditions may affect their use of the target language. For example, in the present study, the researcher's role influenced learners' discourse by allowing a monologue performance only. However, if the learners performed the tasks with each other, their performance may become more dialogic as they may interact with each other in order to negotiate meaning or ask for clarification etc. Nevertheless, it is expected that the target language involving relative clauses would still be produced as the learners would be aware to produce the form during pre-task planning. In conclusion, although socio-cultural theory claims predictions cannot be made regarding the effects of tasks on language use, the results of the present study appear to provide one way of enabling tasks to have generalizable effects on performance. The combination of designing focused tasks that attempt to elicit specific linguistic forms along with the provision of pre-task guided planning that draws learners' attention to the target features helps learners to produce language as intended by the researcher.

9.4.1 Practical implications of guided planning and task complexity for teachers in Japan

The results of this study are of practical use for university teachers in Japan for a number of reasons. First, this study showed how narrative story-telling tasks can be designed using limited financial resources to elicit a grammatical feature known for its

difficulty in L2 oral production with Japanese learners: relative clauses. The study then showed how the tasks can be sequenced together from simple to complex in order to develop learners' accuracy of relative clauses and 3rd person singular and plural, as well as producing gains in fluency. The task sequencing treatment is intended to be manageable for teachers instructing intermediate level university learners because students can effectively plan for the tasks autonomously and then perform the tasks to listening students. As discussed in the last sub-section, intermediate level learners already have declarative knowledge of RCs and can therefore rehearse and practice the target forms without needing teacher assistance or feedback. Thus, this study's task sequencing treatment could be compatible with large class sizes because most students should be able to plan and perform the tasks independent of teacher guidance.

Furthermore, this study's procedure does not require a lot of class time. In terms of duration, this study lasted seven weeks therefore it provides flexibility for teachers to implement into a standard fifteen week university semester that may require additional time for exams and public holidays etc. In addition, the task sequencing treatment is not expected to be time consuming within individual lessons. As we saw in 5.5, the treatment began with a fifteen minute teacher-led guidance session on relative clauses with correct use of 3rd person singular and plural. Students then planned independently for ten minutes before performing one task. In the following week, students performed two tasks with seven minutes planning time per task. The subsequent week involved students performing two tasks with four minutes planning time per task. As a result, each stage of the treatment is not expected to take up the majority of a 90 minute lesson therefore teachers could implement the study without having to compromise too much time needed for other curricular activities.

So why would university teachers in Japan implement this task sequencing treatment into their lessons? The results of this study showed how a series of focused tasks provided practice opportunities for Japanese students to independently develop their use of relative clauses as well as improving fluency. The stable results of the delayed post-tests also showed that guided planning and task complexity provided long-term accuracy gains in Japanese learners' use of RCs, as well as gains in fluency. This bodes well within a Japanese educational context given what was discussed at the start of this study in 1.2, as the main aim of MEXT has been to improve Japanese university learners' use of English for communication. The results of this study provide one way of using and sequencing oral tasks to facilitate developments in Japanese learners' L2

communication skills in terms of fluency, accuracy and complexity. Furthermore, the results of this study also provided gains in students' receptive knowledge of relative clauses as well as gains in their oral development, as shown in the results of the grammatical judgement tests in 6.2.4. Thus, teachers could this task sequencing treatment to improve intermediate learners' receptive skills of RCs as well as their oral skills.

So far we have discussed how this study provides a procedure for university teachers to follow to improve Japanese learners' use of relative clauses and 3rd person singular and plural in order to produce phrases such as 'He thinks he likes the dog which has long hair'. However, teachers could use this task sequencing treatment to focus on other aspects of English language as well. For example, in the present study, students narrated the stories in the present tense, as they were instructed to begin each story by saying 'Today,…..'. Consequently, teachers could use the tasks to practice past tense forms as well by instructing students to begin each narration by saying 'Yesterday….' In doing so, learners could practice past tense forms involving the target language, for example, 'He thought he liked the dog which had long hair'. Alternatively, different linguistic forms could also be targeted and practiced such as determiners, for example, use of possessives 'his', 'her' or use of articles 'a' and 'the' as in the following example, '*His* brother thinks he likes *the* dog which has long hair'. Finally, teachers could also focus on lexis related to the topic of the study such as adjectives, for example, 'long', 'short' as well as mental state verbs in order to describe people thinking, for example, 'He thinks.., he believes.., he wonders..'.

9.4.2 Pedagogic implications of the GP and the GUP group's planning strategies

There are a number of pedagogic implications regarding the findings of the GP and GUP group's planning strategies. First, the post-task interviews of the present study were able to report the patterns that emerged as learners planned for oral narratives that increased in complexity over time. The results of this study showed that both groups were willing to follow guided planning instructions and primarily focused on form even though it did not agree with their personal communicative orientations towards L2 speaking. This shows the influence that planning conditions have on Japanese

intermediate university learners. If the planning conditions are guided towards specific linguistic forms that are intended to help learners complete a task, then there is more certainty that learners will attend to form. The present study showed that both groups appeared to acknowledge the value of the grammar guidance provided at week two as they explicitly attended to the forms throughout the treatment regardless of their personal orientation towards L2 communication. However, if learners are allowed to plan independently, as in the GUP group's planning conditions during weeks three and four, this may cause them to adopt additional form-in-meaning strategies during subsequent task versions. These unguided planning conditions tended to facilitate planning strategies that were more inline with learners' natural orientation towards L2 speech.

The findings of this study showed that continuous guided planning, as in the case of the GP group, resulted in significant improvements in L2 accuracy, complexity and fluency which leads us to suggest that attention to specific forms useful for task completion can benefit all three aspects of L2 speech accordingly. Opportunities to practice key grammatical phrases during strategic planning as learners attempt more complex tasks facilitates proceduralisation of the target language. This treatment appears to be a more powerful tool for L2 oral development than a combination of guided and unguided planning as although in the latter case, learners engage in more varied planning strategies, it does not benefit their oral development as much as continuous guided planning and task complexity.

9.5 Limitations

A number of limitations exist regarding the quantitative and qualitative findings of this study. Regarding quantitative limitations, i.e. the effects of guided planning and task complexity on L2 oral development, over-generalisations cannot be made due to the relatively small sample size of the study (n = 26). The majority of previous planning studies have used larger samples sizes, for example, Bygate (2001b) used 48 learners and Yuan & Ellis (2003) used 42 learners. Although the present study has the advantage of tracking learners L2 oral development over time, the implications of the study are limited as just 13 learners were placed in each group. Nevertheless, as we pointed out in 5.2.1, providing we suggest that the findings of this study provide indications about

second year Japanese university learners of English then we can argue that the findings of this study can be generalised to similar populations of students.

Furthermore, as pointed out in 6.2, the small sample size of the study permitted the removal of only a small number of outliers that were not representative of the sample as a whole. As a result, a small number of remaining outliers contained values which skewed the overall results of the accuracy and complexity variables. Subsequently, many statistical analyses could not be performed on SPSS as it assumes data is normally distributed. Unfortunately, with the accuracy and complexity variables, this was not the case.

Another limitation of this study relates to proficiency. The participants of this study were second year Japanese university learners of English who were recruited from intermediate level classes and above and were considered to be of B2 oral level proficiency. Consequently, this study is unable to confirm whether guided planning and task complexity could produce significant gains in L2 oral development with lower level learners. Ellis (2009a) points out that the majority of planning studies to date have focused on approximately intermediate level learners therefore more research needs to be carried out on beginner level learners.

A further limitation relates to the context of the study. This study was carried out in an experimental setting outside of regular class time and was not part of a course program. Although this study reported significant improvements in L2 oral production, it is limited to the context in which the learning took place. For example, other external factors may have influenced the results of the study such as exposure to the target language during regular class time. As with many previous task-based studies that were also carried out in laboratory settings, there is a need for future studies to be implemented within course programs, as Bygate et al. (2009) note:

> the TBLT enterprise will not be able to rely on individual case studies of learners conducted outside the context of programs of instruction, or on laboratory studies, nor on studies carried out in host classrooms in which the use of tasks is investigated without relating their use to the teaching of the ongoing program. Such work provides a valuable contribution in a sense it might be seen as a form of piloting for the empirical grounding of TBLT. However, more widespread pedagogically contextualized research is clearly needed. (p. 497)

In terms of the interviews used in the main study, the findings are limited for two reasons: conducting the interviews in the L2 and the short time duration. Although the

participants were recruited from intermediate level classes and above, and were considered to be of B2 oral proficiency, they generally did not provide much detail regarding their planning strategies. In Ortega's (2005) study, the post-task interviews were conducted in the L1 which no doubt allowed the participants to explain their planning strategies in detail. However, as the present study was unable to conduct the post-task interviews in Japanese, there was no other option but to rely on the students' use of English in order to elicit their planning strategies. Although this procedure was successful with the participants in the pilot study, the participants in the main study were not enrolled in a bilingual university, did not have English TOEFL scores and had much less exposure using English on university campus. Consequently, they were not accustomed to performing interviews in English and this had adverse affects in terms of their ability to respond in detail to the questions asked. Most of the participants only provided brief responses regarding their planning strategies which limited the findings in terms of finding out why learners planned the way they did (see appendix V for the interview transcriptions). In addition, the 6 minute time limit imposed on the interviews in order to collect data from all the participants within the designated time schedules further compromised the quality of the interviews. Only a limited amount of data could be obtained per student which prevented the use of additional questions and further probing in order to obtain more information regarding the learners' planning strategies.

9.6 Future research

The findings of this study have produced a number of possibilities for further research. To begin with, as this study targeted B2 oral level learners, this prompts the question: to what extent does guided planning and task complexity promote L2 oral development with beginner level Japanese learners of English?

The present study investigated the use of specific linguistic forms, specifically, RC types and 3rd person singular and plural. The findings of this study showed that guided planning and task complexity produced significant developments in learners' use of the forms. As there is a lack of longitudinal task planning studies that have examined form-focused instruction, there is a need for future guided planning studies to explore the development of other linguistic forms known for their difficulty in oral L2 production.

Finally, an important area for future research concerns the effects of guided planning and task complexity within university course programs. As this study produced significant gains in terms of fluency, accuracy and complexity for Japanese university learners of English in an experimental setting, future research is warranted to examine how effective guided planning and task complexity could be within an Asian university English course program. Although this may be a challenge, it is hoped the findings of this study will encourage future investigations into the role of guided planning and task complexity.

10. BIBLIOGRAPHY

Anderson, J. (2000). *Learning and Memory: An Integrated Approach.* New York: John Wiley and Sons.

Baralt, M. (2010). *Task complexity, the Cognition Hypothesis and interaction in CMC and FTF environments.* (Unpublished doctoral dissertation). Georgetown University, Washington, DC. USA.

Browne, C., & Kikuchi, K. (2009). English education policy for high schools in Japan: ideals vs reality. *RELC Journal*, 40, 2, 172-191.

Brumfit, C. (1984). *Communicative Methodology in Language Teaching*. Cambridge: Cambridge University Press.

Bulte, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (pp. 21-47). Amsterdam: John Benjamins.

Bygate, M. (2001a). Speaking. In R. Carter, & D. Nunan (Eds.), *The Cambridge Guide to Teaching English to Speakers of Other Languages* (pp. 14-21). Cambridge: Cambridge University Press.

Bygate, M. (2001b). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks. Second Language Learning and Testing* (pp. 23-49). Harlow: Longman.

Bygate, M., Norris, J., & Van den Branden, K. (2009). Understanding TBLT at the interface between research and pedagogy. In K. Van den Branden, M. Bygate, & J. Norris (Eds.), *Task-based language teaching: A reader* (pp. 496–499). Amsterdam: John Benjamins.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task repetition. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 37-77). Amsterdam: John Benjamins.

Bygate, M., Skehan, P., & Swain, M. (2001). Introduction. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks. Second Language Learning and Testing* (pp. 1-21). Harlow: Longman.

Byrne, D. (1987). *Techniques for classroom interaction.* London: Longman.

Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English*. Cambridge: Cambridge University Press.

CLAN Programs from CHILDES. Retrieved from http://childes.psy.cmu.edu

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, L., Manion.L., & Morrison, K. (2005). *Research Methods in Education*. New York: Routledge Falmer.

Creswell, J. (1994) *Research design: qualitative and quantitative approaches.* London: Sage.

Coughlan, P. & Duff, P. A. (1994). Same task, different activities: Analysis of a SLA task from an activity theory perspective. In J. Lantolf, & G. Appel (Eds.), *Vygotskian approaches to second language research*. (pp. 173-194). Norwood, N.J.: Ablex.

De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics,* 13, 1, 1-24.

Dekeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practising second language grammar. In C. Doughty, & J. Williams (Eds.), *Focus on form in Classroom Second Language Acquisition.* (pp. 42-64). Cambridge: Cambridge University Press.

Dekeyser, R. (2003). Implicit and explicit learning. In. C.Doughty, & M. Long (Eds.), *Handbook of Second Language Acquisition*. (pp. 313-348). Malden, Mass: Blackwell.

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.

Dornyei, Z. (2007). *Research Methods in Applied Linguistics.* Oxford: Oxford University Press.

Doughty, C. (1991). Second language instruction does make a difference. *Studies in Second Language Acquisition*, 13, 431-469.

Douglas, D. (2001). Performance consistency in second language acquisition and language testing research: a conceptual gap. *Second Language Research*, 17, 4, 442-456.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

Ellis, R. (2005a). Planning and task-based performance: Theory and research. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language.* (pp. 3-37). Amsterdam: John Benjamins.

Ellis, R. (2005b). Measuring implicit and explicit knowledge of a second language: a psychometric study. *Studies in Second Language Acquisition*, 27, 141-172.

Ellis, R. (2008). *The Study of Second Language Acquisition (Second Edition)*. Oxford: Oxford University Press.

Ellis, R. (2009a). The effects of three types of task planning on the Fluency, Complexity, and Accuracy in L2 Oral Production. *Applied Linguistics*, 30, 4, 474–509.

Ellis, R. (2009b). Task-based language teaching: sorting out the misunderstandings. *International Journal of Applied Lingustics*, 19, 3, 221-246.

Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press.

Erkman, F, R., Bell, L., & Nelson, D. (1988). On the Generalization of Relative Clause Instruction in the Acquisition of English as a Second Language. *Applied Linguistics*, 9, 1, 1-20.

Eysenck, M. (2001). *Principles of Cognitive Psychology (Second Edition)*. Hove: Psychology Press.

Field, A. (2009). *Discovering Statistics using SPSS (Third Edition).* London: Sage.

Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85-102). New York, New York: Academic Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in second language acquisition*, 18, 3, 299–323.

Foster, P., & Skehan, P. (1999). The influence of source planning and focus of planning on task-based performance. *Language Teaching Research*, 3, 3, 215-247.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 3, 354–375.

Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Eds.), *Perspectives on Fluency* (pp. 243-265). Ann Arbor: University of Michigan Press.

Gass, S, M., & Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research.* Mahwah, N.J.: Lawrence Erlbaum.

Gilabert, R. (2007a). *Task Complexity and L2 Narrative Oral Production*. (Unpublished Doctoral Dissertation). University of Barcelona, Spain.

Gilabert, R. (2007b). The simultaneous manipulation of task complexity along planning time and [+- Here-and-Now]. In M.D.P. Garcia-Mayo (Eds.), *Investigating tasks in formal language learning* (pp. 44-68). Clevedon, UK: Multilingual Matters.

Hakuta, K. (1976). A case study of a Japanese girl learning English as a second language. *Language Learning*, 26, 2, 321-351.

Hamilton, R. (1994). Is implicational generalisation unidirectional and maximal? Evidence from relativization instruction in a second language. *Language Learning*, 44, 1, 123-157.

Housen, A., Kuiken, F., & Vedder, I. (2012), Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency* (pp. 1-21). Amsterdam: John Benjamins.

Howatt, A. (1984). *A History of English Language Teaching*. Oxford: Oxford University Press.

Howell, P. (2004). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Europe PMC Funders Group Author Manuscripts*, 31, 1-30.

Izumi, S. (2003). Processing difficulty in comprehension and production of relative clauses by learners of English as a second language. *Language learning*, 53, 2, 285–323.

Izumi, Y., & Izumi, S. (2004). Investigating the Effects of Oral Output on the Learning of Relative Clauses in English: Issues in the Psycholinguistic Requirements for Effective Output Tasks. *The Canadian Modern Language Review*, 60, 5, 587-609.

Johnson, K. (2001). *An introduction to Foreign Language Learning and Teaching*. Harlow: Pearson Education Limited.

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 143-165). Amsterdam: John Benjamins.

Keenan, E., & Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Enquiry*, 8, 1, 63-99.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P.Robinson (Eds.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 39-61). Amsterdam: John Benjamins.

Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145-164.

Krashen, S., & Scarcella, R. (1978). On routines and patterns in second language acquisition and performance. Language Learning, 28, 283-300.

Krashen, S., & Terrell, T. (1983). *The Natural Approach: Language Acquisition in the Classroom.* Oxford: Pergamon.

Kuno, S. (1974). The position of relative clauses and conjunctions. *Linguistic Inquiry*, 5, 1, 117-136.

Lazaraton, A. (2005). Quantitative research methods. In E. Hinkel (Eds.), *Handbook of Research in Second Language Teaching and Learning* (pp. 209-224). Mahwah, N.J.: Lawrence Erlbaum.

Lennon, P. (2000). The Lexical Element in Spoken Second Language Fluency. In H. Riggenbach (Eds.), *Perspectives on Fluency* (pp. 25-42). Ann Arbor: University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: Cambridge University Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.

Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (pp. 171-199). Amsterdam: John Benjamins.

Long, H, M., & Robinson, P. (1998). Focus on form; Theory, research, and practice. In C. Doughty, & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 15-42). Cambridge: Cambridge University Press.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes, & S.M. Gass (Eds), *Tasks and Language Learning*: *Integrating theory and practice* (pp. 123–167). Clevedon: Multilingual Matters Limited.

Mackey, A. (1999). Input, interaction and second language development: an empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21, 557–587.

Menard, S. (2002). *Longitudinal Research*. Thousand Oaks, Calif.:Sage.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83-108.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2003). *Regarding the Establishment of an Action Plan to Cultivate "Japanese with English Abilities"*. Retrieved from http://www.mext.go.jp/english/topics/03072801.htm

Mochizuki, N., & Orgtega, L. (2008). Balancing communication and grammar in beginning level foreign classrooms. *Language Teaching Research*, 12, 11, 11-37.

Murphy, R. S. (2013). Critical Review of Five Major L2 Syllabus Designs. *Center of Fundamental Education (Kiban Center) Kiyo, Journal of the University of Kitakyushu*, 15, 169-221.

Murphy, V. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, 13, 1, 34-65.

Norris, J., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. Norris, & L. Ortega (Eds.), *Synthesizing Research on Language Learning and Teaching* (pp. 3-53). Amsterdam: John Benjamins.

Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics*, 30, 4, 555-578.

O'Donnell, K. (2005). Japanese secondary English teachers: Negotiation of educational roles in the face of curricular reform. Language, Culture, and Curriculum, 18, 3, 300-315.

Ohata, K. (2004). Phonological differences between Japanese and English: Several potentially problematic areas of pronunciation for Japanese ESL/EFL learners. *Asian EFL*, 6, 4, 1-19.

Oppenheim, A., N. (2000). *Questionnaire design, interviewing and attitude measurement.* London: Continuum.

Ortega, L. (1995). The effect of planning in L2 Spanish narratives. *Research Note 15*. Honolulu HI. University of Hawai'i, Second Language Teaching and Curriculum Center.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21, 109-148.

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R.Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 77-111). Amsterdam: John Benjamins.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: recent trends and future directions. *Annual review of Applied Linguistics,* 25, 26-45.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30, 4, 590-601.

Prabhu, N. S. (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

Reinders, H., & Cho, M. (2012). Enhancing Informal Language Learning with Mobile Technology – Does it Work? *Journal of Second Language Teaching and Research,* 1, 1, 3-29.

Richards, K. (2003). *Qualitative Inquiry in TESOL*. Basingstoke: Palgrave Macmillan.

Robinson, P. (2001). 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.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, 43, 1–32.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake, and perceptions of task difficulty. *International Review of Applied Linguistics*, 45, 193–213.

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz, & L. Sicola (Eds.), *Inside the Learner's Mind: Cognitive Processing in Second Language Acquisition* (pp. 243-268). Amsterdam: John Benjamins.

Robinson, P. (2011). Task-based language learning: a review of issues. *Language learning*, 61, 1, 1–36.

Robinson, P., & Gilabert, R. (2007). Introduction: Task Complexity, the Cognition Hypothesis, second language learning and performance. *International Review of Applied Linguistics*, 45, 3, 161-177.

Robson, C. (2002). *Real World Research*. Oxford: Blackwell.

Sakui, K. (2004). Wearing to pairs of shoes: language teaching in Japan. *ELT Journal*, 58, 2, 155-163.

Sakui, K. (2007). Classroom management in Japanese EFL classrooms. *JALT Journal*, 29, 1, 41-58.

Samuda, V., & Bygate, M. (2008). *Tasks in Second Language Learning*. Basingstoke: Palgrave Macmillan.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 111-143). Amsterdam: John Benjamins.

Schachter, J. (1974). An error in error analysis. *Language Learning*, 24, 2, 205–214.

Schmidt, N. (2010). *Researching Vocabulary*. *A Vocabulary Research Manual*. London: Palgrave MacMillan.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 4, 357-385.

Seedhouse, P. (1999). Task-based Interaction, *ELTJ*, 53, 3, 149-156.

Seedhouse, P. (2005). "Task" as Research Construct. *Language Learning*, 55, 3, 533-570.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 1, 93-120.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning. The influence of surprise information and task time on second language performance. In R.Ellis (Eds.), *Planning and Task performance in a second language* (pp. 193-219). Amsterdam: John Benjamins.

Sorace, A. (1996). The use of acceptability judgements in second language acquisition research. In W. C. Richie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375–405). New York: Academic Press.

Swan, M. (2005). Legislation by Hypothesis: The Case of Task-Based Instruction. *Applied Linguistics*, 26, 3, 376-401.

Swan, M., & Smith, B. (2001). *Learner English*. Cambridge: Cambridge University Press.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Eds.), *Planning and Task Performance in a Second Language* (pp. 239-277). Amsterdam: John Benjamins.

Tavakoli, P., & Foster, P. (2008). Task Design and Second Language Performance: The Effect of Narrative Type on Learner Output. *Language Learning*, 58, 2, 439-473.

The Common European Framework of References for Languages. (2010). Retrieved from http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf.

Thompson, C., & Millington, N. (2012). Task-based learning for Communication and Grammar Use. *Language Education in Asia*, 3, 2, 159-167.

Thompson, C., & Jones, C. (2013). Aims and Outcomes. *The Asian EFL Journal*, 66, 28-42.

Timmis, I. (2003). *Corpora, classroom and context: the place of spoken grammar in English language teaching.* (Unpublished doctoral dissertation). University of Nottingham, UK. Retrieved from http://etheses.nottingham.ac.uk

Willis, J. (1996). *A Framework for Task-Based Learning.* Harlow: Longman.

Willis, D., & Willis. J. (2001). Task-based language learning. In R. Carter, & D. Nunan (Eds.), *The Cambridge Guide to Teaching English to Speakers of Other Languages* (pp. 173-180). Cambridge: Cambridge University Press.

Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly*, 26, 4, 693-711.

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.

11. APPENDICES

Appendix A: The Triadic Componential Framework for task classification (Robinson & Gilbert, 2007, p. 164)

Task complexity (Cognitive factors)	**Task conditions (Interactive factors)**	**Task difficulty (Learner factors)**
(Classification criteria: cognitive demands)	(Classification procedure: interactional demands)	(Classification criteria: ability requirements)
(Classification procedure: Information-theoretic analyses)	(Classification procedure: behaviour-descriptive analyses)	(Classification procedure: ability assessment analyses)
a. resource-directing variables making cognitive/conceptual demands	**a. Participation variables** making interactional demands	**a. Ability variables** and task-relevant resource differentials
+/- here and now	+/- open solution	h/l working memory
+/- few elements	+/- one-way flow	h/l reasoning
+/- spatial reasoning	+/- convergent solution	h/l task switching
+/- causal reasoning	+/- few participants	h/l aptitude
+/- intentional reasoning	+/- few contributions needed	h/l field independence
+/- perspective-taking	+/- negotiation not needed	h/l mind/intention-reading
b. resource-dispersing variables making performative/procedural demands	b. Participant variables making interactant demands	**b. Affective variables** and task-relevant state-trait differentials
+/- planning time	+/- same proficiency	h/l openness to experience
+/- single task	+/- same gender	h/l control of emotion
+/- task structure	+/- familiar	h/l task motivation
+/- few steps	+/- shared content knowledge	h/l processing anxiety
+/- independency of steps	+/- equal status and role	h/l willingness to communicate
+/- prior knowledge	+/- shared cultural knowledge	h/l self-efficacy

Appendix B: student participation form (pilot study)

AGREEMENT TO PARTICIPATE IN PILOT STUDY

1 June 2011

Name of teacher/researcher: Colin Thompson

Name of learner: _____

Date: _____

I consent to Colin Thompson using any data I give for purposes of his research study. The data will include the following: interview responses regarding planning strategies for speaking tasks, pre-speaking test and post-speaking test results, speech performance of speaking tasks that have been audio recorded, student questionnaires regarding planning strategies for speaking tasks. I agree to let this data be viewed by the researcher and his research supervisor as part of an ongoing research project. I am aware that I have the right to withdraw from the study at any point of the research process and I understand that if I do so, all data relating to me will be destroyed. I also understand that any data I do provide will be used only for the intended purposes and will be anonomised so I cannot be identified from the data.

Finally, I am aware that the results of the data will be written about as part of the research project and that details of the results may be published in academic journals or discussed at teaching conferences.

Signed (by learner)

……

THANK YOU VERY MUCH FOR YOUR CO-OPERATION

Appendix C: Seven obligatory cases of relative clauses used in Mochizuki & Ortega's (2008) task (Mochizuki & Ortega, 2008, p. 35)

Context 1 (OS): I like the dog which has long ears

Context 2 (OO): I want the dog which the little girl has in her arms

Context 3 (OS): I like the dog which has long hair

Context 4 (OO): I want the dog which many people are watching

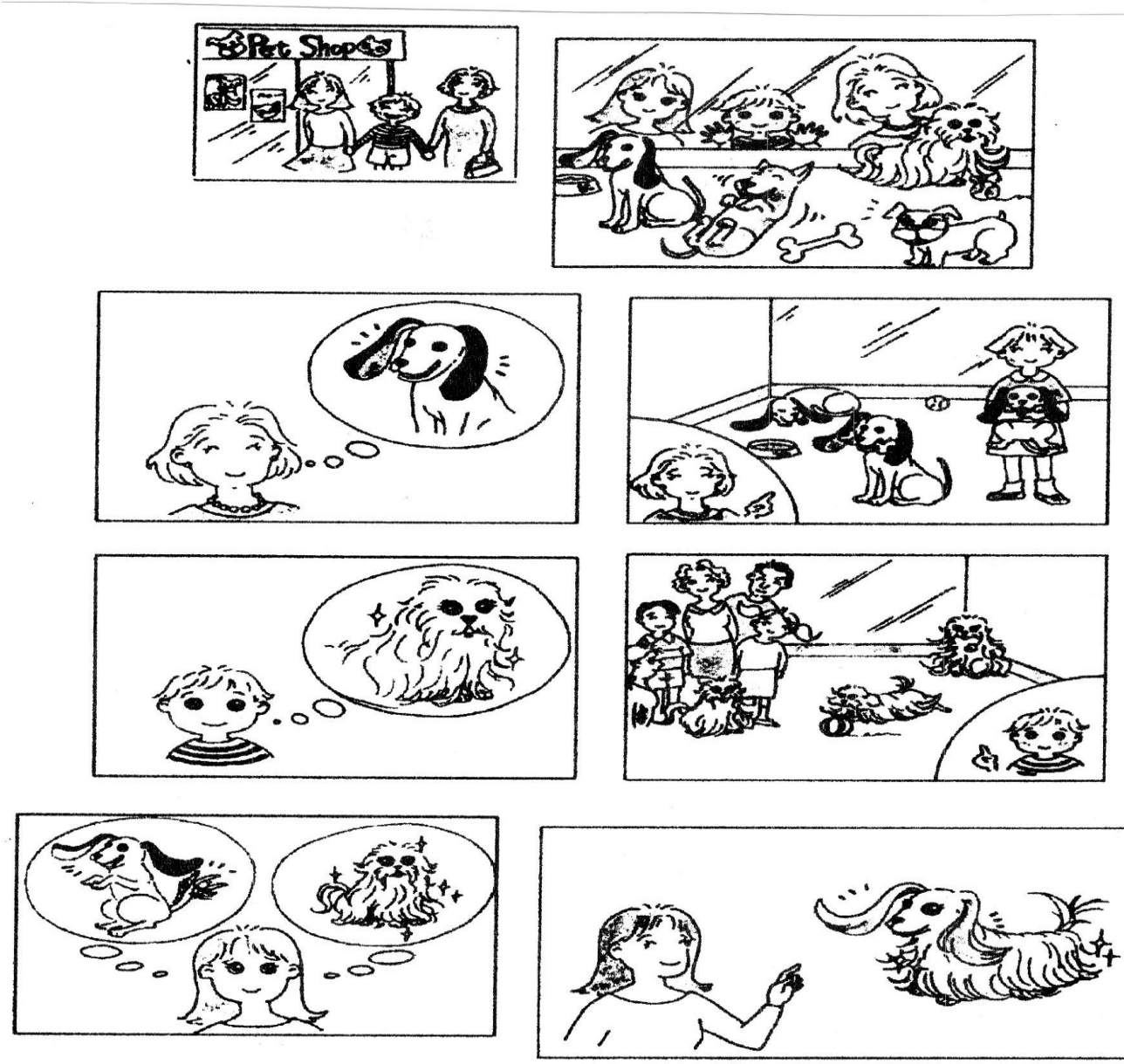
Context 5 (SS): The dog which has long ears looks friendly

Context 6 (SS): The dog which has long hair is beautiful

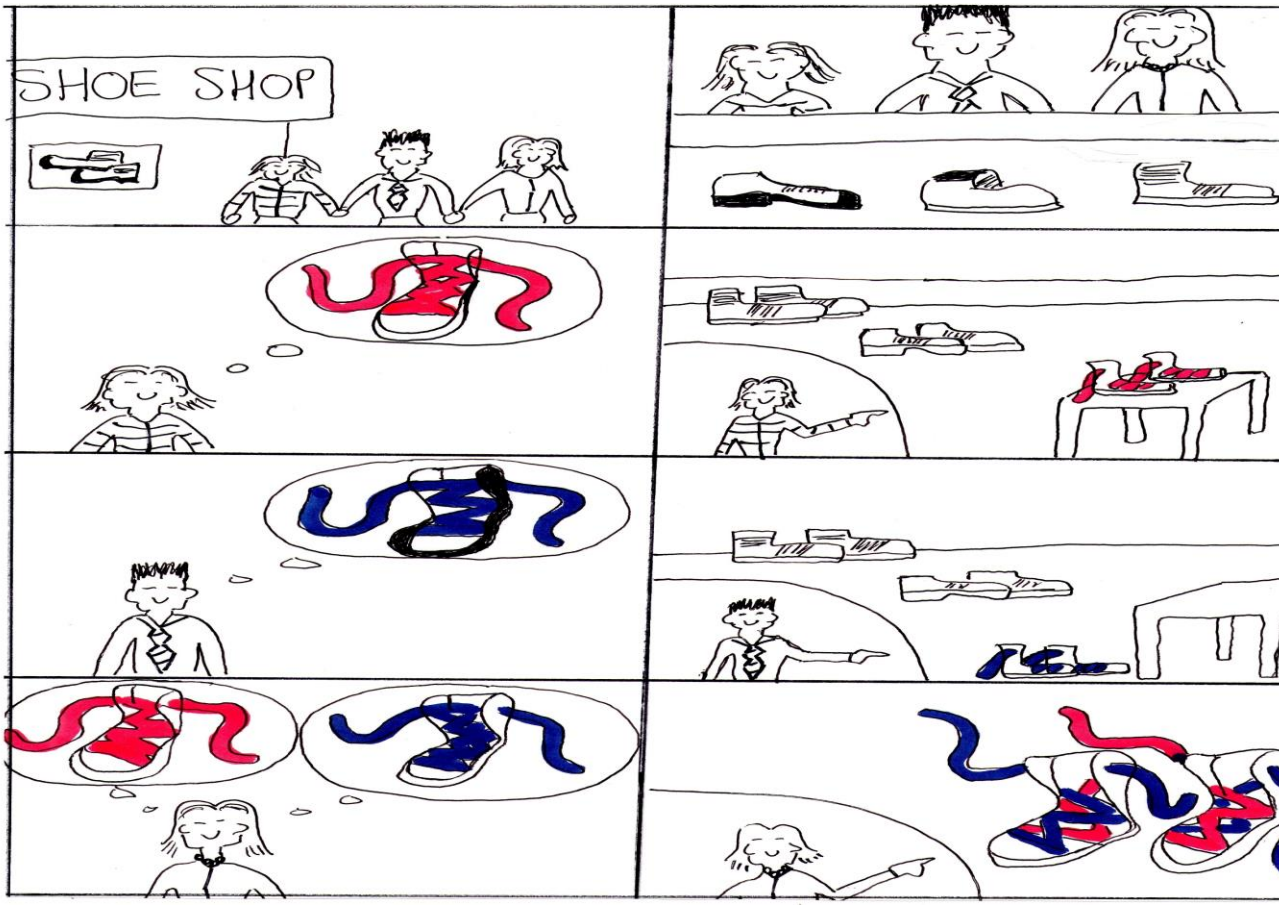
Context 7 (OO): I want the dog which has long ears and long hair

Appendix D: Pre-and post-test narratives (pilot study)

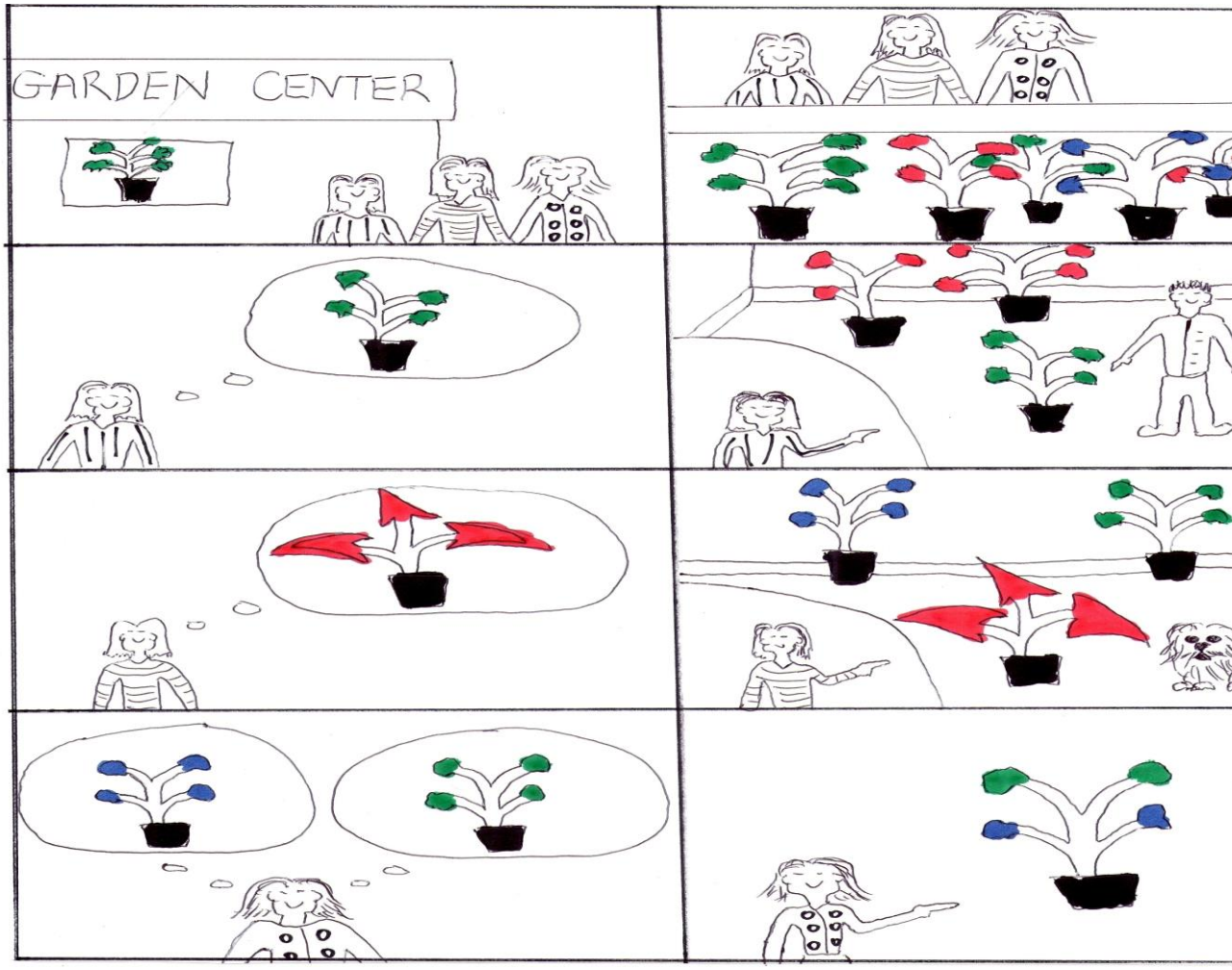
Pre-test narrative (Mochizuki & Ortega, 2008, p. 36)



Immediate post-test narrative (pilot study)

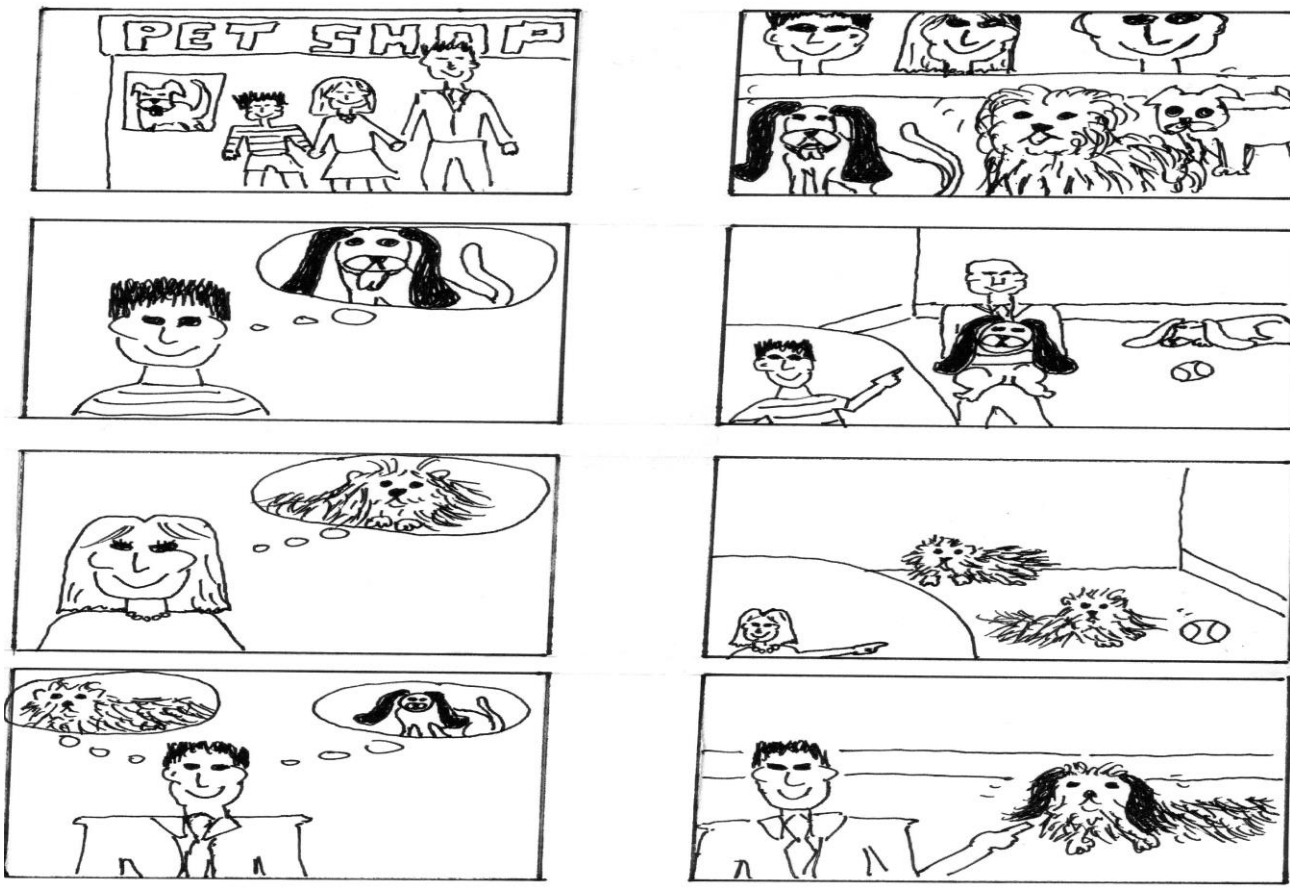


Delayed post-test narrative (pilot study)

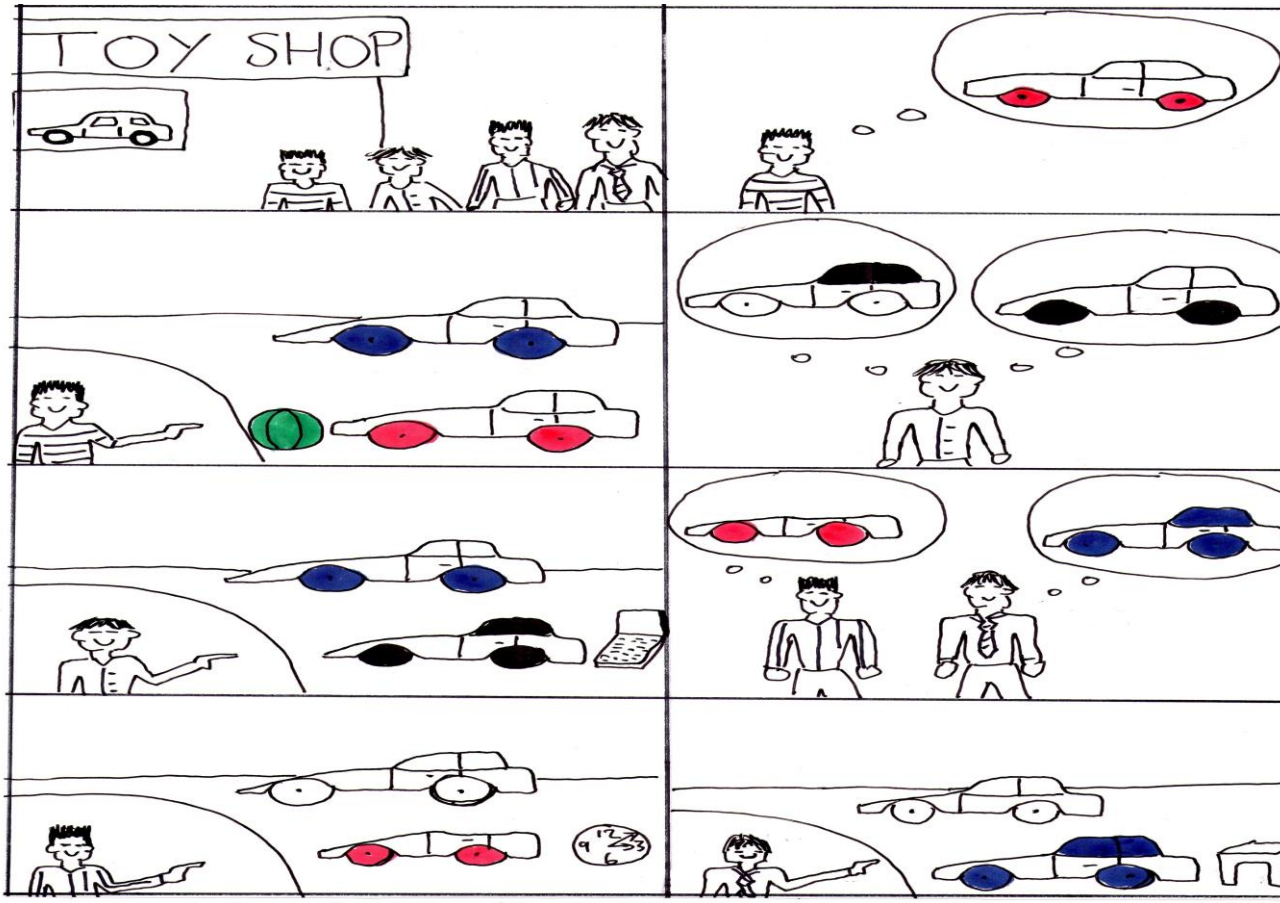


Appendix E: Treatment narratives (pilot study)

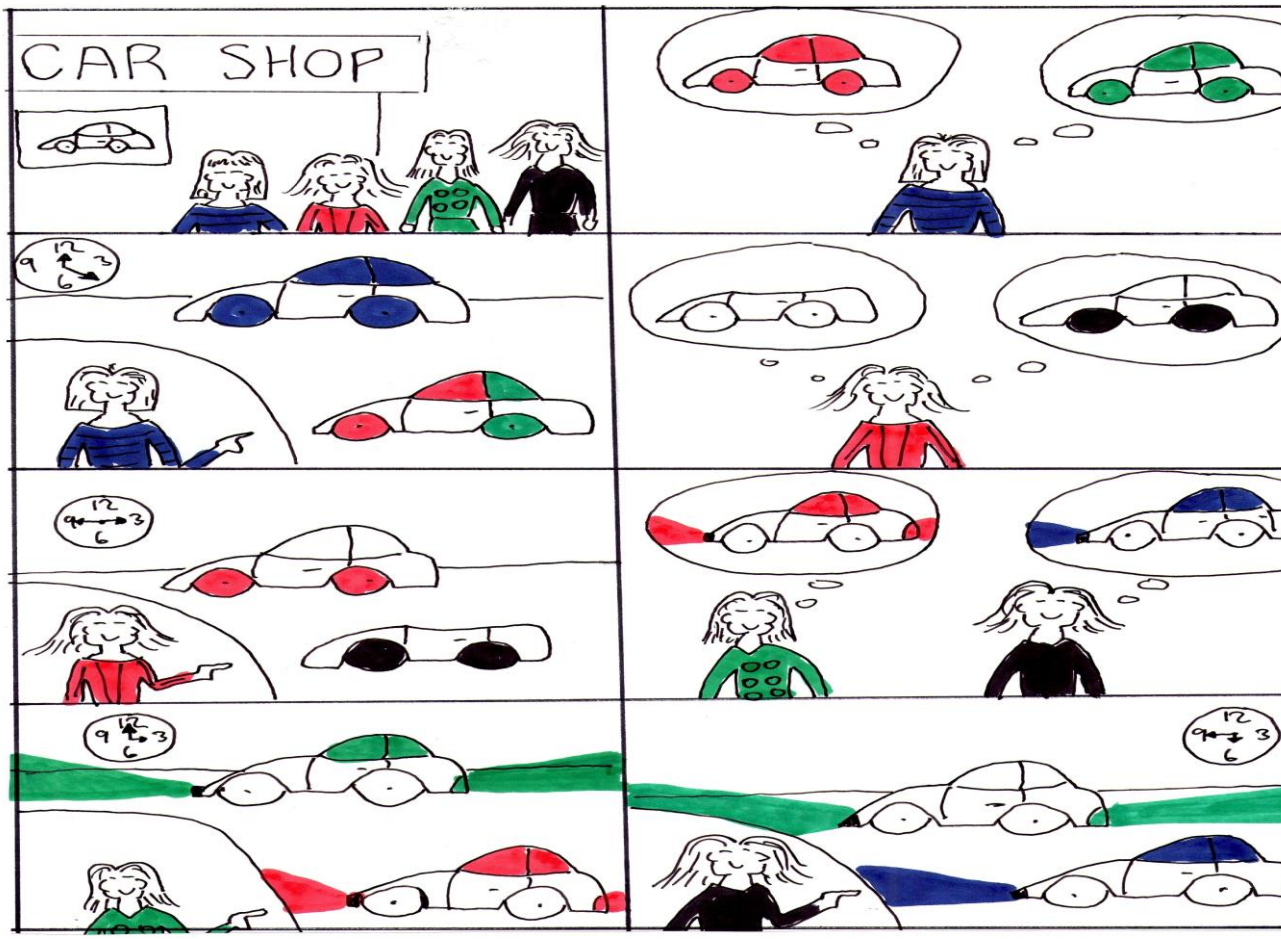
Narrative one



Treatment narrative three (pilot study)



Treatment narrative five (pilot study)



Appendix F: Pilot study guided planning note-sheets (adapted from Mochizuki & Ortega, 2008; Sangarun, 2005, p. 133)

Name_____

1. Think about your partner's knowledge of the pictures.

2. Describe each picture separately.

3. Think of transition words such as *first, next*

4. Think about grammar when planning. For example

She thinks that she likes the dog which looks friendly

She wants the doll which has black shoes

She believes that she likes the doll which has short hair

She wants the doll which the girl is watching

5. Make notes in English but don't write everything in detail. Try and talk without your notes.

Appendix G: Pilot study post-task interview questions (adapted from Ortega, 1999, p. 148; Yuan & Ellis, 2003)

1. Ice-breaker

How was that?

2. Opening

'Did you plan before the start of the task?'

'How did you plan for it?'

'What would you say your focus was when you prepared the story?'

'Did you think about grammar, vocabulary, how to organize your story, or something else?'

'What would you say your focus was when you were doing the task?'

Reactive questions:

What do you mean?

Why?

Can you give me an example?

Can you say that again?

Can you explain that a little bit more?

When did you have time to think of that? / When were you thinking that?

'In what way?'

What was difficult about it?

Repeated task

1. Ice-breaker

How was that?

'Did you plan before you started the task?'

'How did you plan?'

'Did you plan differently compared to last time?' 'In what way?'

'Did you write a lot?'

'What would you say your focus was when you prepared the story?'

‘Did you think about grammar, vocabulary, how to organize your story, or something else?’

‘Did you feel pressured at any time?’ ‘Compared to last time?’

Guided planners only

‘How useful were the note sheets?’

‘How useful were the grammar explanations?’

Appendix H: Pre- post-test narrative transcriptions (pilot study)

B1 Intermediate Guided Planner: B1GP

B1 Intermediate Unguided Planner: B1UP

Pre-test: B1GP

err one day err kevin his sister and mother went to pet shop
err err then err a err kevin's mother err want the dog that has big ear
err but kevin wants the dog very err err much hair
but kevin's sis ter wants both
then she decided to choose err the dog err has big ears and much hair

Pre-test: B1UP

err one day kevin and her err his his sister kate and his mother went to pet shop
and they saw many kinds of dogs
long hair short ear long ear and little dog
and mo ther his mother like long ear one err and mother like mother want to err mother
err kevin err mother want kevin to play with that kind of dog
err but kevin like a dog which has long hair
and he want to play with long hair dog
and her sister like like liked long haired and long ear and long tail
so his sister want long hair err err wanted a dog which has long ear and long tail and beautiful long long hair

Immediate post-test: B1GP

err to day peter and his brother o sister err go to sh shoe shop
err peter and his brother and his sister err found three shoe
shoe shoes in the shop
first peter peter thinks that he likes the shoes which has red laces
then peter wants the shoes which has red laces on the chair
second peter's brother thinks that he likes the shoes which

has err blue laces
then peter's brother wants the shoes which has err blue laces
err under the chair
finally peter's sister wants the shoes which has red and blue laces
then peter's sister err wants the shoes which has the blue and red laces
finish

Immediate post-test: B1UP

today peter and his brother and his
err his sister are at the shoe shoe shop
err there are a lot err various kind of shoes
and peter wanted to buy sneaker which has red sh red shoe lace
and peter find find at the store peter fou found the shoes
like that on the tab chair on the on the table
peter grabbed to see that to find that
and his brother wanted to buy a a sneaker which has blue
shoe lace and black heel so sole
and at the shop he find the shoe like that
but it has white heel and it's by the table
and his sister want to buy two kinds of shoes
one of them has re re red shoe lace and
another the left has blue shoe lace
at the shop his sister find the shoes which has blue and red shoe laces
she is glad to find that and she decided to buy it

Delayed post-test: B1GP

today err kate err with her two sisters at the garden center
kate and her sisters err look for the plant
with flowers that they want
first kate thinks that she likes the plant with green flowers
then kate wants the plant with green flowers next to the man

second kate's second sis ter thinks that she likes the plant
which has the gr err red flowers
then kate's second sister wants the plant which has the green
err red flowers next to the dog
kate's third sister wants the plant which has blue flowers and green flowers
then kate's third sister wants the plant which has green flowers and blue flowers

Delayed post-test: B1UP

today kate and her sister one and sister two are at a garden center
there are many kinds of flowers
kate wanted to buy a buy a plant which has green green flowers
and at the shop kate found a plant su like that
err because because shop tailor shop tailor
introduced introduced that so she decided to buy that
after that kate her sister two wanted to buy a plant
which has red flowers
at the shop err her sis her sister two find find a plant like that that is that is next to the dog
she decided to buy that
after that her sister three wanted to buy two kinds of plants
one of them has blue flowers another has green flowers
at the shop she found a plant which has green flowers and blue
flowers both of them
sh she she she is very interested in it so she decided to buy that
that's all

Appendix I Pilot study student interview transcriptions

B1 Intermediate Guided Planner: B1GP

B1 Intermediate Unguided Planner: B1UP

Week 2: B1GP

Interviewer in bold

How was that speaking task?

A bit tough

A bit tough, ok why was that?

Er because its not very easy to explain in details

Could you?... in what way?

How the look like erm yeah er the like the pictures is very erm complicated for me to explain

OK did you plan before you started the task?

Yeah a little

Ok how did you plan for it?

First I see the pictures and try to find out how the pictures described

Ok could you give an example of that?

Like erm a boy with his parents I think his parents er went to the zoo and looking for the monkeys and and then I described the monkey that has erm like big ears, long arms, hairy like that

Ok now when you were planning what would you say your focus was when you were planning the story?

Focus, erm erm how the monkeys look like

How the monkeys look like? Ok could you explain that it alittle bit more detail?

Like each pictures there are many three different kinds of monkey like that has long arms, and other one has big ears and third one has a er big ears and hairy

Ok erm ok now did you write a lot?

Erm not too much

Not too much ok why was that?

Er I because I couldn’t find out more details from the pictures

Ok so what were you thinking about then when you were writing?

Erm I was thinking erm the relations er between the pictures

Ok so when you were planning, were you thinking about grammar or vocabulary how to organize the story or some thing else?

Er I was thinking about how to organize the story

Ok really what was difficult about it?

Er for me er writing I mean organising the story is easy for me but when going to speak I'm not good at it so I couldn't organize very well

Ok erm did you feel under pressure at any time?

Er Yeah a little yeah

Could you explain why?

Erm always worries that is my expression or explaining is good or bad

Have you ever done anything like this before?

Er yes but not much

Ok how useful were the note sheets?

Sorry?

How useful were the notesheets when you were planning?

Er helped me to form my explanation

In what way?

Er er to organize the story

Ok how useful were the grammar explanations?

Er er it's helpful

It's helpful, could you give me an example?

Erm Kevin thinks that the monkey

Oh no I mean give an example of why it was helpful?

Er erm when there's no erm there's no help with this erm this grammar maybe I er stuck

Ah right so its

Its very hard for me to make the sentences without the grammar.

Without the grammar ok I see.

Week 3: B1GP

How was that speaking task?

Erm so so.

So so ok and and what do you mean by that?

Erm I thought that I can do better before I was doing this task

Oh really?

but when I was doing the speaking I couldn't do it that good way than I thought.

Ok why was that?

Erm maybe erm I didn't have enough time to fix words

Ok and did you plan before you started the task?

Yes I did.

Ok how did you plan?

I tried to follow the rules from the sheet. I tried to make up my own sentences but actually I couldn't

You couldn't, why was that?

because I was very worrying about is my grammar ok.

Ok did you plan differently compared to last time?

Erm yeah yeah.

You did, in what way did you plan differently?

I for the first time I mean before I just looked at the pictures and describe only a little detail but this time I tried to describe more details

Why was that?

because I thought that its not good if I don't describe more details because there are a lot of drawings in the pictures so er yeah

A lot drawings in the pictures. In what way are there a lot of drawings in the pictures?

Erm, drawings

Can you explain that? Can you give me an example?

Like a girl pointing the doll then that there are only only one doll there are another another object but it think this er research before and I didn't describe the other details but this time I tried to describe another like er yeah.

Ok ok did you write a lot?

Er yeah.

Ok why was that?

Well its going to help me to speak

Ok in what way will it help you?

Erm to help me to speak more details erm,

Ok and what is details?

About the drawing from the pictures

Ok the pictures ok what would you say your focus was when you were planning the story?

The person er the way I focused on the action that person does

The actions

Like pointing and what they are thinking and what they are wanting.

Ok and when were you thinking about that? At the start or the end?

At the start.

Ok and when you were planning, did you think about the grammar, the story or vocabulary or something else?

I tried to er I mean I was very conscious about grammar

The grammar why is that?

because if I don't use the correct grammar then my speaking will be very chaos.

Ok erm so compared to last time, how was it different?

This time I'm not very nervous

Ok

I'm very relaxed

And why was that?

Maybe I get accustomed to this work.

Did you feel under pressure at any time?

Yes a bit

A bit, can you give me an example?

I'm not very good at English so when I speak I sometimes get stuck what I'm going to say.

Ok when was that?

When, sometimes.

Sometimes ok so did you feel under pressure at any time, now compared to last time,

Did you feel under pressure, compared to last time?

Ok because?

It's better, yeah less pressure because I get accustomed I got the confident.

And why did you get more confident?

Because the first time I did this work I didn't know what exactly what to do or how to make it better or something like this, I just didn't have erm idea to do erm.

Ok well how how useful were the notesheets?

Note sheets erm its helpful

Helpful, in what way?

er er help me to describe the pictures

Ok and how useful were the grammar expressions, how useful was that this time?

Er it's a one of a base to make the sentences yeah,

Ok I see so so how useful was it?

Er its useful erm

Useful compared to last time?

Yeah I think this time is more useful erm

Ok any example why? Why it it more useful this time?

Because first time I er I'm not get accustomised this work not sure I mean I didn't er I do know how to use it but to make up the story and organise it its very difficult for me to make the sentences.

Week 4: B1GP

How was that speaking task?

Er so so.

So so, ok, ok in what way was it so so, so so alittle easy, a little difficult, in what way was it a little difficult, a little easy?

Erm I tried to describe all the details which during the picture but I couldn't make it and just I couldn't finish

You couldn't finish, that's why it was difficult?

For example, I think there was a clock in the picture but I had no idea how to describe in words

Ok, alright, ok did you plan before you started the task?

Yeah,

How did you plan?

Yeah first I look at the pictures and describe what to say and I tried to say what the picture written er drawing

Ok did you plan differently compared to last time?

Erm, as I said, erm I tried to managed to describe all the picture yeah its drew drew

When?

Today

How was that different to last time?

Erm well I think I describe a lot than before

Ok so planning was different to last time because you

Yeah tried to describe more details

More details ok I see was that it? Anything else?

Oh yeah i tried not to say er er er

Oh really? Why was that?

Because maybe maybe it doesn't feel so well

Ok ok alright now did you write a lot when you were planning?

Er not this time,

Ok why was that?

I was very puzzled to describe whole things

Why was that?

Because there are a lot of details in the pictures and I tried to fix it and decided what to say but not enough time

Not enough time ok I see so when you were planning, what would you say your focus was when you were planning?

You mean this time? er focus, er more details

That's it?

And not try to say some er er

Ok ok now when you were planning, did you think about grammar, vocabulary, how to organise the story or something else?

Er organize story I mean main focus more details

Ok more details, I see, more details does that mean more grammar or organising the story?

Its organizing the story

Organising the story ok did you feel under pressure at any time?

A little,

Ok can you explain that in a little more detail?

Erm I was a bit nervous this time if I stop speaking what should I do, messed up or confused

Compared to last time, any different?

I think today is better

Today is easier?

Not easy but better

So feeling under pressure compared to last time, less pressure? Same?

Er same

Any reason why? Same pressure

Today erm task was abit difficult

More difficult?

More difficult

Why was it more difficult?

Because I has more details than before

Ok how useful were the notesheets this time?

This time I use it but not very much

Not as much as last time? Why was that?

Well I tried to write as much as possible but I had no time for fixing ideas

Ok, ok how useful were the grammar explanations this time?

It's very helpful to help fix the stories

Week 2: B1UP

How was that for you?

Er I have a lot of mistake

Can you give an example, in what way?

first I said by mistake I said past tense

ok can you explain that alittle bit more?

Just tense and vocabulary?

Yeah

How did you plan? What did you do?

erm first I think that situation in Japanese

ok

and translate into English

ok I see

I know that is not good but

Can you explain that abit more?

I have to translate in English

Were you thinking like that all the time?

yeah it takes very long time more time than other people

ok so what was difficult about it?

not I can't speak naturally I can't use these words

ok I see ok so what would say your focus was when you were preparing the story?

What they're doing focus their face smiling or

You were looking at the pictures?

Yeah

Ok Did you write a lot?

Write about yeah er no not so much

Ok anything difficult about the writing?

I'm not good at using which or that I don't know how to say

Ok erm did you think about grammar or vocabulary or how to organise the story or something else?

Er organize the story.

Ah really Ok did you think about organise the story all the time or something else?

How about grammar or vocabulary?

Grammar is difficult.

Ok so were were planning to organize the story? How did you prepare organzing the sotry when you wre planning?

I tried to explain erm I tried to use same words and same grammar same grammar.

Ok and when wre you thinking that? At the start or All the time?

Yeah I tried but I couldn't

Were you thinking like that at the start or all the time?

All the time.

Thinking the same way? Ok

Ok so did you feel under pressure?

Yeah

Ok in what way?

Yes I had to speak not fast and pronounce clearly

I see i see have you done anything like this before?

Er no

Week 3: B1UP

How was that speaking task?

I was confused because I couldn't understand what the picture mean.

Ok and what do you mean by that? Can you give me an example?

I can couldn't make a story

Ok why?

Erm I couldn't couldn't understand I couldn't connect the pictures

Ok ok so did you plan before the start of the task?

Yes I did,

Ok how did you prepare for it?

Er you said don't write it completely but I wrote I wrote details sorry

Ok that's ok now why did you do that?

I was worried

Ok and could you explain that a little bit more?

Ah I can't speak er fluently I can't respond

Ok ok so er did you prepare differently compared to last week?

No I didn't similar I think

Similar ok similar in what way?

Not way I but I wrote a lot of detail and I wrote sentences

Ok so last week and this week you were writing a lot ok I see ok so did you write a lot?

Yes I did

Ok so this time what would you say your focus was when you were preparing the story?

Focus was er sisters wanted to buy similar dolls so I I focus the word the same as

Ok I see erm so this time when you were planning, did you think about grammar, vocabulary, how to organise the story or something else?

Er erm past tense yeah erm

Ok and when were you thinking that, at the start or the end or all the time?

All time I was confused because you said to me please start speaking the story kate and her sister one, two, three are at the but I said she wanted

Ok

Or she decided

Why was that?

Kate want and find and decides and all of them same past tense

Erm same tense

Erm is unnatural

Unnatural

I though

Ok ok did you feel under pressure at any time?

Yeah

Can you give an example of why you were under pressure?

Er I wrote in detail on this paper so I can't see this paper little wanted to see this paper.

Ok I see and compared to last time were you under pressure? Same same feeling?

More

More?

Because this difficult

Why why was this difficult compared to last time?

I feel erm last time I could find difference easily and I could make a story more easily than this time

Because?

Because I this dolls is her their friends

Ok

So I was confused and panic.

Week 4: B1UP

How was that speaking task?

Maybe I got used to it this time

Ok in what way did you get used to it?

How erm er similar

Similar?

Every time the picture

Can you give me an example?

Every time I use the same grammar

Ok I see, did you plan before the start of the task?

Yes

How did you plan?

Erm er same same as last time yeah,

Same as last time? for example?

I saw the picture and I write I write in English what they are

Ok now, did you plan different compared to last time?

Erm different

Did you plan different compared to last time?

No always similar I used same grammar almost

Almost all the same? Ok?

All ah but this time there are clocks

Clocks yeah?

Pictures of clocks so I said about it

Ok did you write a lot?

No less than last time

Why was that?

Because I got used to it that type of task I can speak English easier

Ok, in what way could you speak English easier?

I can remember what types of grammar I should use

Ok, erm ok so what would say your focus was when you were planning the story?

Focus was er every girls girls what they want want erm

Ok can you give an example of that?

Example er erm erm I explain more detail details and example kate want to find the cars, one of them the other I explain almost completely

Ok, erm right so when you were planning, did you think about grammar or vocabulary or something else?

Once more

Ok, erm right so when you were planning, did you think about grammar or vocabulary or something else?

Er

Yeah so you were planning for ten minutes, did you think about grammar or vocabulary or something else?

I yeah I thought about vocabulary not vocabulary grammar

Grammar

But I didn't think about vocabulary

Ok and can you explain that a little bit more?

And this time I wanted to I wanted to speak more like a story so I said I add the last sentence like a concluding sentence.

I see I see ok alright did you feel under pressure at any time?

Yes but decreasing

Decreasing and why is that?

I getting used to this type of task

Ok

And I have a little confidence

I see ok so compared to last time, did you feel under pressure the same way?

Less less yes

Less because you're getting more confident? And in what way are you getting more confident?

Yeah and I when I can't speak smoothly I confused and get panic panic but this time I'm getting used to it so I can speak more fluently than last time.

Appendix J: Main study participant consent form

AGREEMENT TO PARTICIPATE IN RESEARCH STUDY

4th June 2012

Name of researcher: Colin Thompson

Name of learner: _____

Date: _____

I consent to Colin Thompson using any data I give for purposes of his research study. My participation will involve 6 hours of paid English classes that will include the following: interview responses regarding planning strategies for English speaking tasks, English speaking test results, speech performance of speaking tasks that have been audio recorded, student questionnaires regarding the difficulty of speaking tasks.

I have agreed to be paid 5,000Y for 6 hours of English participation (800Y per hour) for Colin's Thompson's research study. I agree to let this data be viewed by the researcher and his research supervisor as part of an ongoing research project. I am aware that I have the right to withdraw from the study at any point of the research process and I understand that if I do so, all data relating to me will be destroyed. I also understand that any data I do provide will be used only for the intended purposes and will be anonomised so I cannot be identified from the data. Finally, I am aware that the results of the data will be written about as part of the research project and that details of the results may be published in academic journals or discussed at teaching conferences.

Signed (by learner)

……

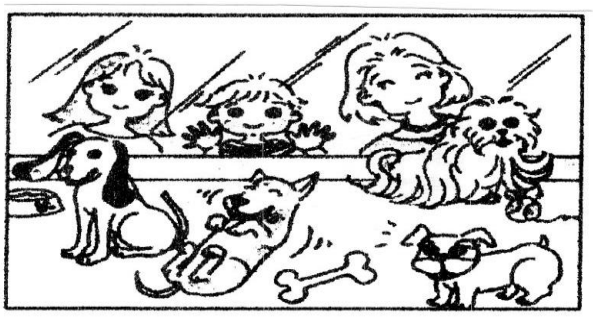
THANK YOU VERY MUCH FOR YOUR CO-OPERATION

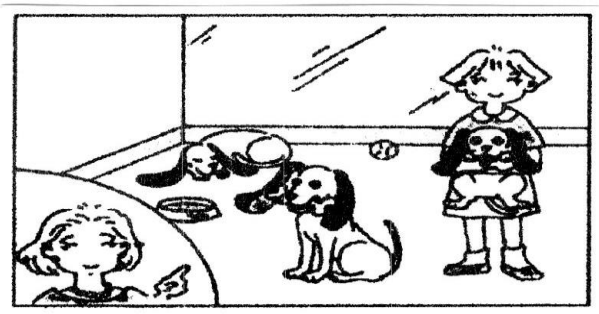
Colin Thompson

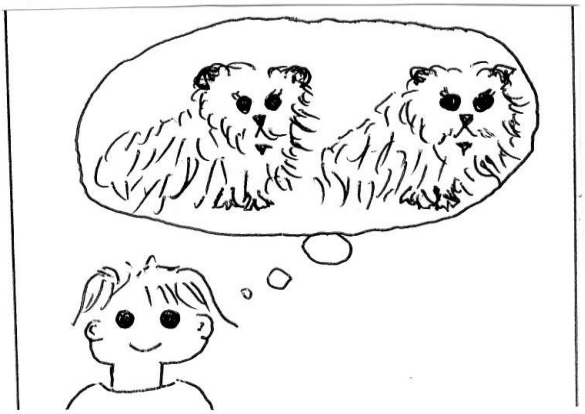
6-17 Hatsune Machi, Tobata-ku, Kitakyushu, Fukuoka, 804-0066, Telephone: 08039047885

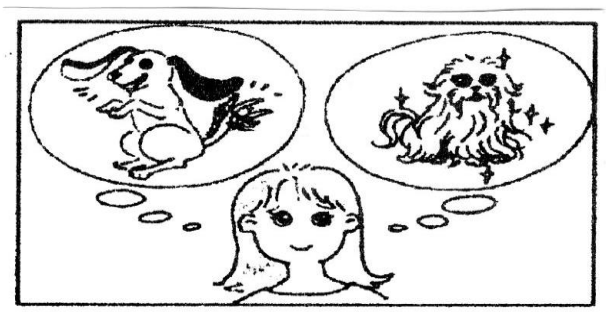
Appendix K: Pre- and post-test narratives (main study)

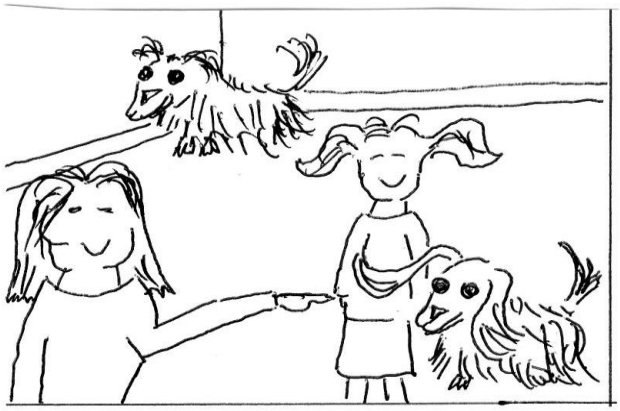
Pre-test narrative (adapted from Mochizuki & Ortega, 2008)



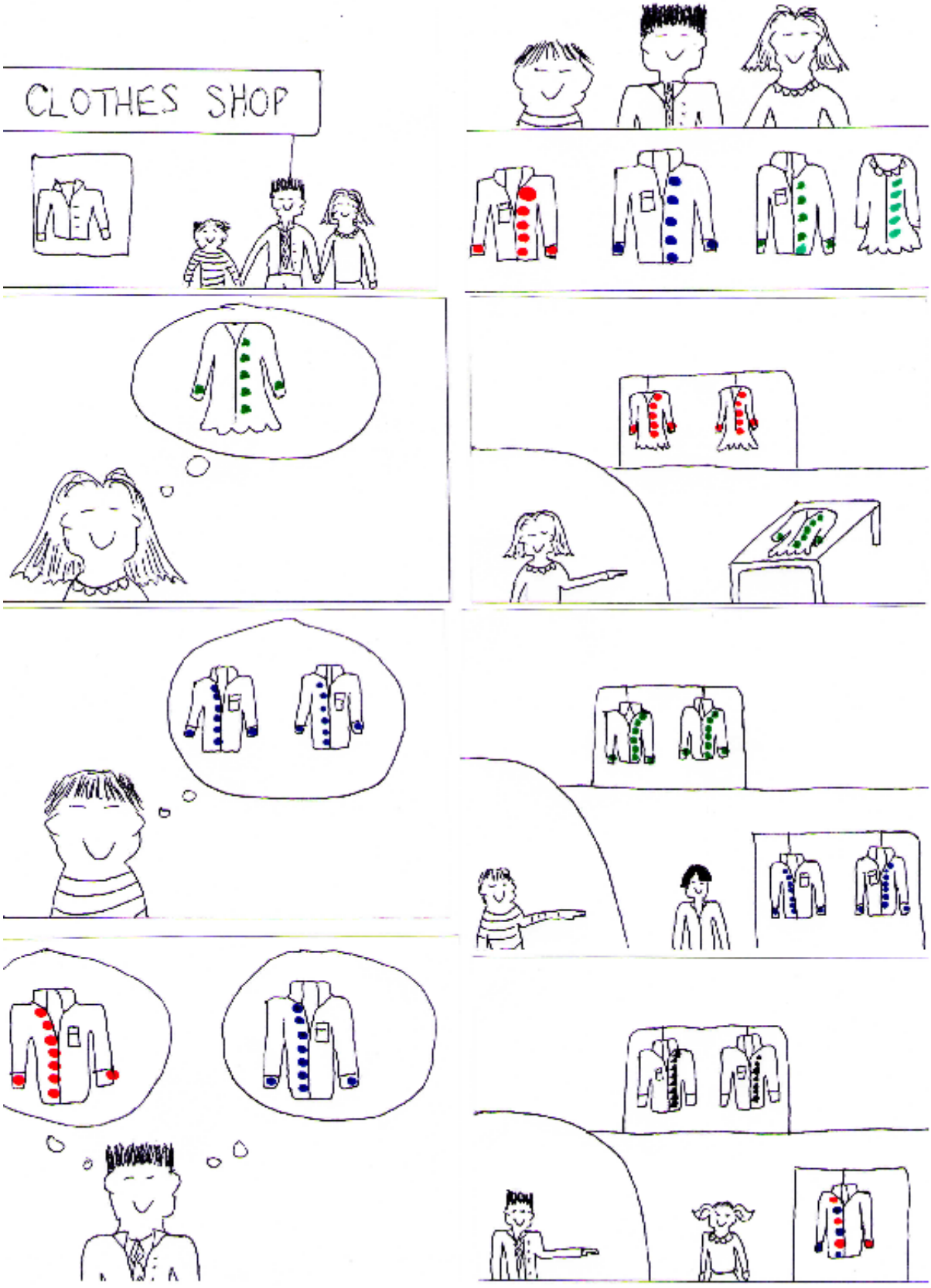




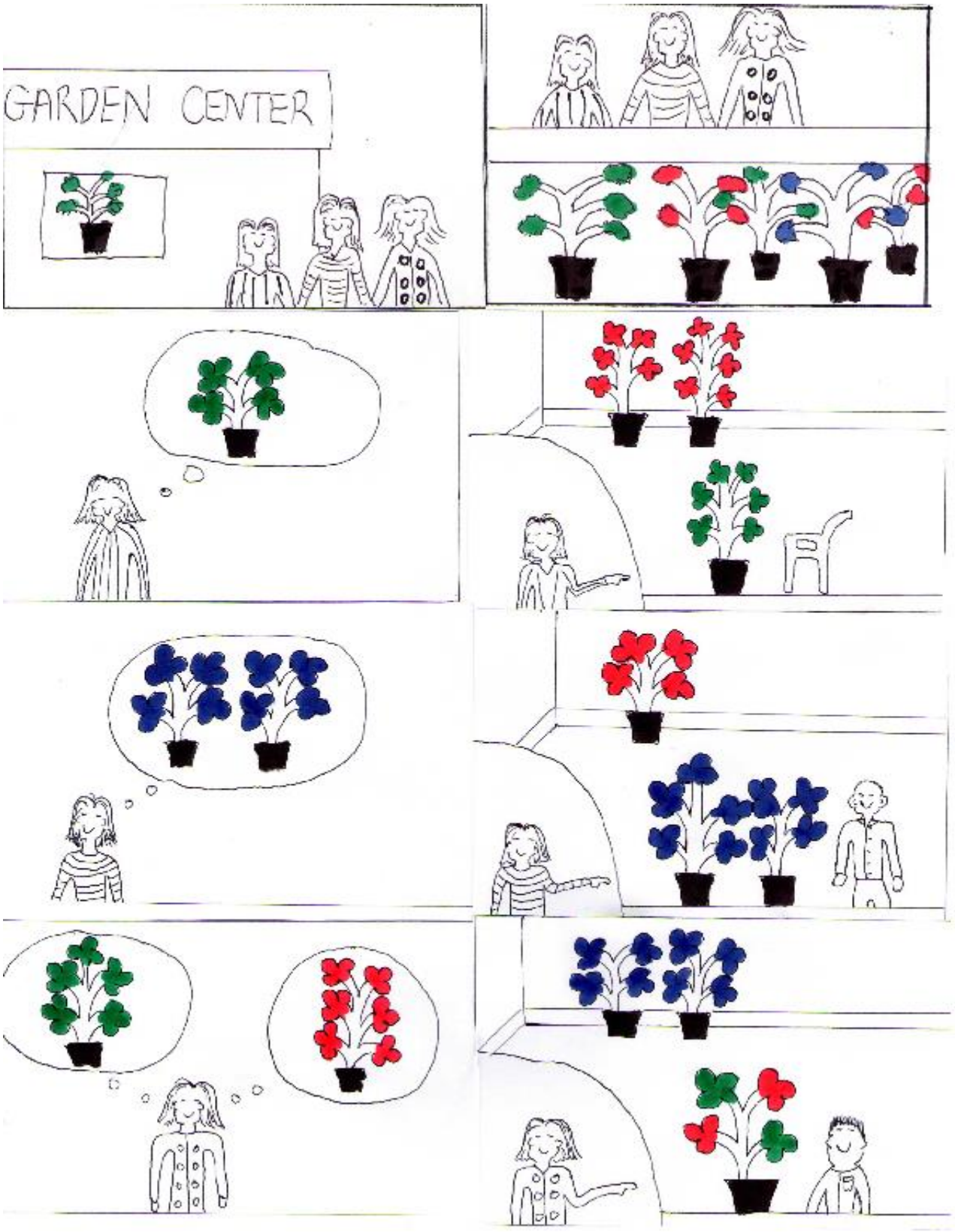




Immediate post-test (main study)



Delayed post-test (main study)



Appendix L: Grammatical Judgement test

Immediate post-test (adapted from Izumi, 2003; Reinders & Cho, 2012)

Grammar Test

Name:_____

The book who Peter read was given to James.

_____

_____

The car drove too slow.

_____

_____

The boy that he was dancing with Kate was the one I was smiling at.

_____

_____

The man who you play soccer with is sick.

_____

_____

He wants the cats which has white hair.

_____

_____

We want the cake which the girl is looking at.

_____

_____

Bill think that he likes the woman who is dancing.

_____

_____

In was cold and it snowed heavy.

_____

_____

The cat has long hair which likes James.

_____

_____

Ken read the book which was given to Harry.

_____

_____

The girl who was dancing with Peter was the one I was talking to.

_____

_____

The woman who is old likes John.

_____

_____

In the mountains, you have to walk very carefully.

_____

_____

James believe that he played a piano which was made in England.

_____

_____

John think that he eats oranges which are grown in France.

_____

_____

The dog likes the food who Chris made.

_____

_____

Every player in this team has to practice tonight.

_____

_____

He met who the man went to the hospital.

_____

_____

Neil likes the rabbit which John is playing with.

_____

_____

He spoke so quickly that nobody understood him.

_____

_____

The boy who was in pain saw the doctor.

_____

_____

Bob likes the dog which Jim it is playing with.

_____

_____

The book which Kate read was given to Mary.

_____

_____

All students should study hardly.

_____

_____

The woman who Sally saw was the one we were talking about.

_____

_____

The man who you met went to the hospital.

_____

_____

The boy who you met him went to the hospital.

_____

_____

He finished his meal very quickly.

_____

_____

John believes that he wears shirts that have no buttons.

_____

_____

The student you study with who is absent.

_____

_____

We ate the food the boys which were looking at.

_____

_____

Simon read the paper was given to James.

_____

_____

He completed his study successful.

_____

_____

She met the woman who went to the hospital.

_____

_____

He wants the car which have small wheels.

_____

_____

The horse which has long hair is sleeping.

_____

_____

I looked for the book which Sally was talking about.

_____

_____

He is famous because he sings beautiful.

_____

_____

The boy which was in pain saw the doctor.

_____

_____

She thinks that she likes the shirts which have pink buttons.

_____

_____

The ice-cream which was eaten it by Harry was the one we were looking at.

_____

_____

The cat which have long hair is sleeping.

_____

_____

He looked for the book who Tom was talking about.

_____

_____

He thinks that he likes the woman who has long hair.

_____

_____

Ken bought puppies which has short hair.

_____

_____

The cake which was eaten by Mary was the one we were looking at.

_____

_____

The cat likes the food that Paul made.

_____

_____

The woman Bill who saw was the one we were talking about.

_____

_____

He has learned English for three years, but speaks poorly.

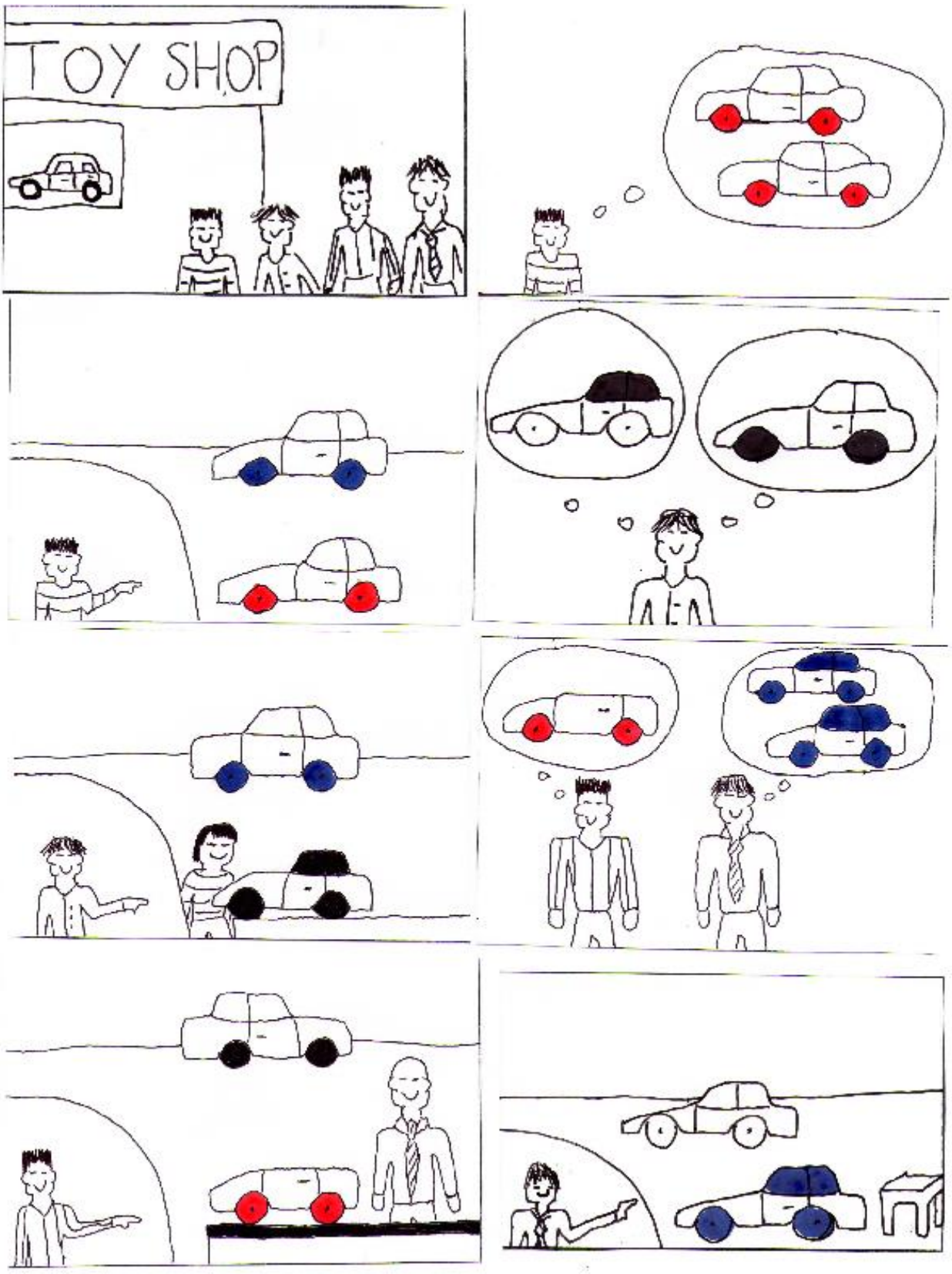
_____

_____

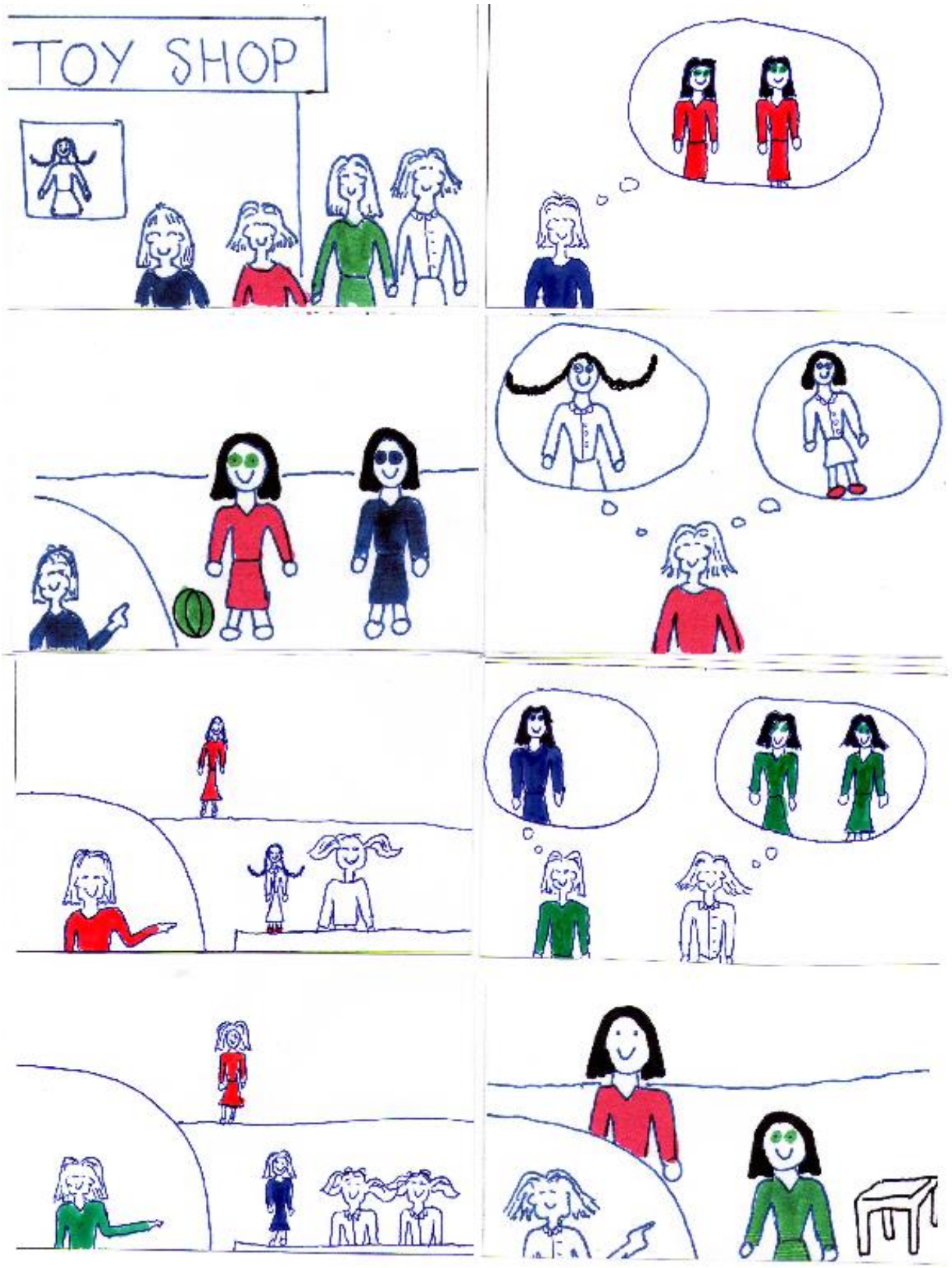
Appendix M: Treatment tasks (main study)

Treatment task one

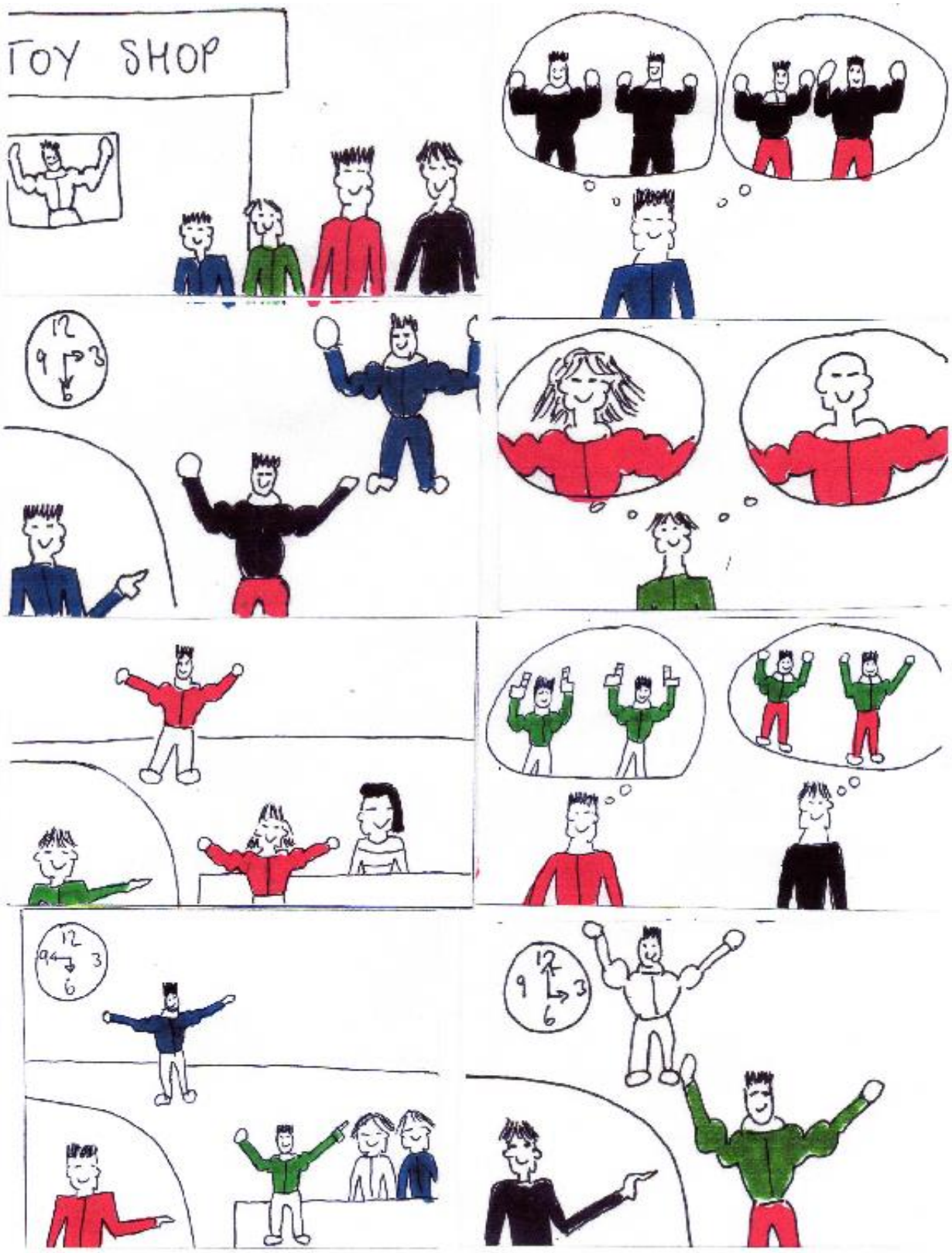
Treatment task two (main study)



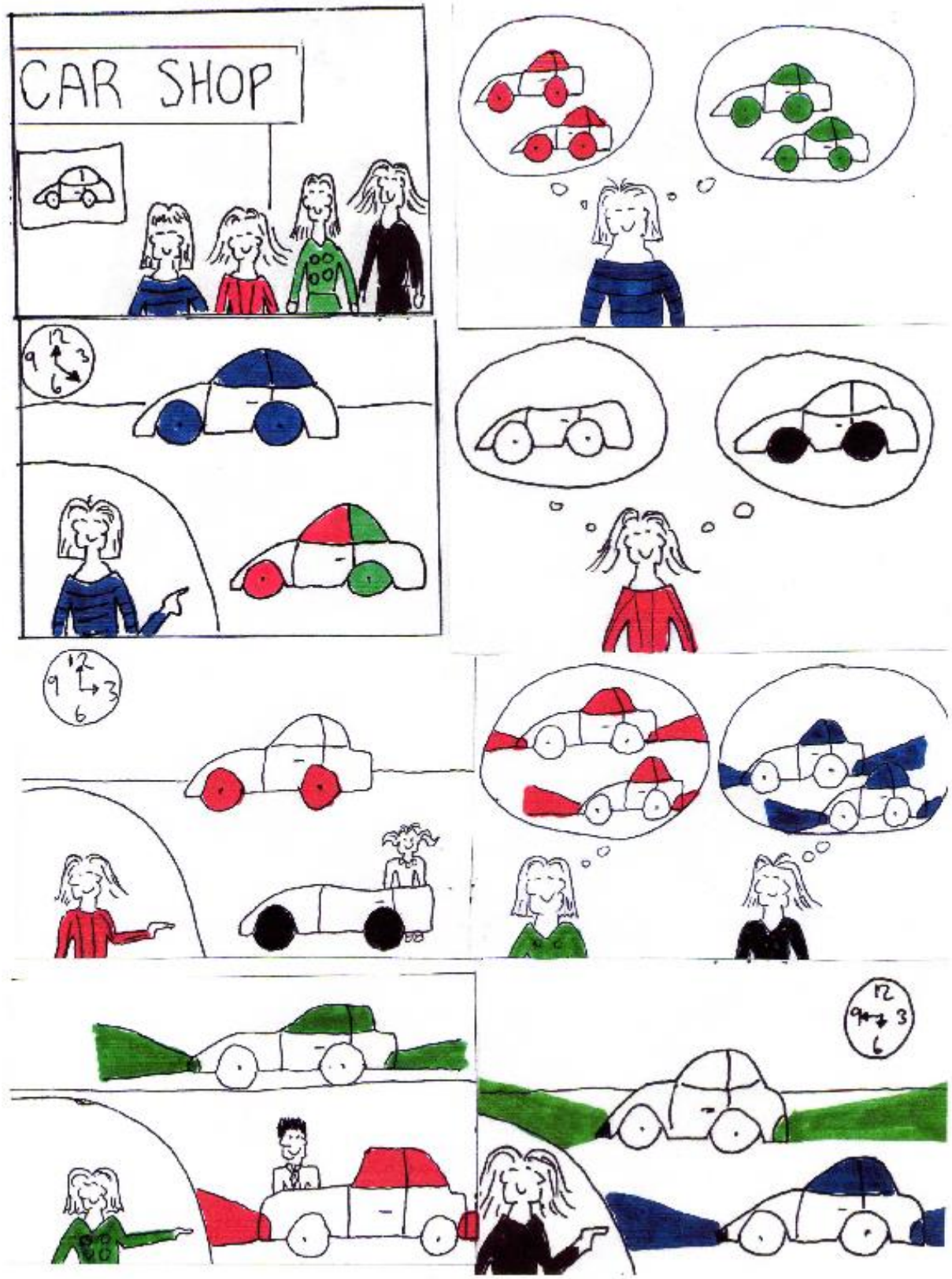
Treatment task three (main study)



Treatment task four (main study)



Treatment task five (main study)



Appendix N: Guided planning note-sheet (GP group) main study

Name_____

1. Try to use this grammar when planning. For example:

Kate thinks that she likes the cat which has blue eyes

She wants the cat which the teacher is looking at

James believes that he likes the cat which has a long tail

He wants the cat which the teacher is talking about

Appendix O: Main study post-task interview questions (adapted from Ortega, 1999, p. 148; Yuan & Ellis, 2003)

1. Ice-breaker

How was that?

2. Opening

'Did you plan before the start of the task?'

'How did you plan for it?'

'What would you say your focus was when you prepared the story?'

'Did you think about grammar, vocabulary, how to organize your story, or something else?'

'What would you say your focus was when you were doing the task?'

Reactive questions:

What do you mean?

Why?

Can you give me an example?

Can you explain that a little bit more?

'In what way?'

What was difficult about it?

Repeated task

1. Ice-breaker

How was that?

'Did you plan before you started the task?'

'How did you plan?'

'Did you plan differently compared to last time?' 'In what way?'

'Did you write a lot?'

'What would you say your focus was when you prepared the story?'

'Did you think about grammar, vocabulary, how to organize your story, or something else?'

'Did you feel pressured at any time?' 'Compared to last time?'

Appendix P: Pre-treatment Questionnaire

Please tell me how much you agree or disagree with the following statements by circling a number from 1 to 7:

Strongly disagree	Disagree	Slightly disagree	Unsure	Slightly Agree	Agree	Strongly agree
1	2	3	4	5	6	7

1. "*I often think about the errors I make when speaking English and sometimes I don't want to speak if I make a mistake. I'm never happy with my grammar."*

1 2 3 4 5 6 7

2. "*I think communication is more important than grammar. When learning to speak English, it's natural to make mistakes but through practice we can improve."*

1 2 3 4 5 6 7

Thank you!

Appendix Q: Pre- post-test narrative transcriptions (main study)

Guided Planner: 112

Guided and Unguided Planner: 214

Pre-test: 112

today kevin his mother and his sister at pet shop
they are looking they are looking dogs
his mother his mother want a a dog having a black ear
err kev kevin wants kevin wants two dogs <having>
having as err
his sister his sister want also them
but she decided to buy to buy a dog kevin wanted

Pre-test: 214

kevin went to the pet shop with her mother and sister
they watched many dogs
and mother want to long ear dog
and kevin played with them
but kevin want to different dogs which my mother want to buy
and kevin's sister wants to buy two dogs which my mother and Kevin want

Immediate post-test: 112

today tim and his family are at clothes shop
they are looking at some shirts and dress
his mother thinks that she likes dresses err
dress which has blue err green buttons
she wants the shirts which is on the table
tim thinks that he likes shirts which have blue buttons
he wants the shirts which he wants the shirts which the man
is looking at

his mother err his father thinks that he likes shirts which have red
buttons and blue buttons
he want he wants the shirts which the girl is looking at

Immediate post-test: 214

today tim and his father and his mother at clothes shop
err they they are watching red buttons shirts and blue
buttons shirts and green buttons shirts and green buttons dress
his mother wants want wants the dress which has green buttons
she thinks that he wants the dress on which is on the table
and his father wants the two shirts which have blue buttons
he thinks that he wants the shirts which are near by sailor
and tim wants two shirts which has red buttons and which has
blue buttons
he thinks that he wants the shirts which has red and blue buttons
and which is near by a girl

Delayed post-test: 112

today sam and her sisters are at the garden center
they are looking plants
sam thinks that she likes plant which have which has green leave
she wanted she wants the plant which which is behind the chair
first sister thinks that she likes plants which have blue leave
she wants the plants which the man is looking at
sam's sister thinks that she likes plants which have which has green leave or red red leave
she thinks that she err she wants the plant which has she
she wants the plant which the boy is looking at

Delayed post-test: 214

today sam and her sisters at garden center
they watching many flowers
sam wanted flower which which has green flower
and she thinks that she wants the flower which is nearby chair
and her sister who wear bordershirts wants two flowers which
have blue blue flowers
she think that she wants two flowers nearby a man
and her sister who wear dot dot shirts want two flowers
which has green flowers and red flowers
she think that she wants the flower which is nearby a man

Appendix R: Instructions for coding and calculating accuracy (Main study)

Only seven obligatory RC contexts are assessed per narrative test thus providing a maximum total score of 60 points (see examples below).

Pre-test narrative: possible examples for maximum points for each relative clause context

Context 1 (OS): The mother thinks that she likes the dog which has long ears	9
Context 2 (OS): She wants the dog which is next to the girl	8
Context 3 (OS): Kevin thinks he likes the dogs which have long hair	9
Content 4 (OPREP): He wants the dog which the family is looking at	8
Context 5 (OS): Kate thinks she likes the dog which has long hair	9
Context 6 (OS): She also thinks she likes the dog which has long ears	9
Context 7 (OPREP): She wants the dog which the girl is smiling at	8
Maximum total:	60

Immediate post-test narrative: possible examples for maximum points for each relative clause context

Context 1 (OS): Tim's mother thinks that she likes the dress which has green buttons	9
Context 2 (OS): She wants the dress which is on the table	8
Context 3 (OS): Tim thinks he likes the shirts which have blue buttons	9
Content 4 (OPREP): He wants the shirts which the man is looking at	8
Context 5 (OS): Tim's father thinks he likes the shirt which has blue buttons	9
Context 6 (OS): He also thinks he likes the shirt which has red buttons	9
Context 7 (OPREP): He wants the shirt which the girl is looking at	8
Maximum total:	60

Delayed post-test narrative: possible examples for maximum points for each relative clause context

Context 1 (OS): Sam thinks that she likes the plant which has green leaves	9
Context 2 (OS): She wants the plant which is next to the chair	8
Context 3 (OS): Sam's sister thinks she likes the plants which have blue leaves	9
Content 4 (OPREP): She wants the plants which the man is looking at	8
Context 5 (OS): Sam's other sister thinks she likes the plant which has red leaves	9
Context 6 (OS): She also thinks she likes the plant which has green leaves	9
Context 7 (OPREP): She wants the plant which the boy is looking at	8
Maximum total:	60

Instructions for coding the seven obligatory contexts

Use the rating scale below, code each of the seven contexts per narrative. Depending on the quality of the context produced, each student may receive a grade from 0 – 9 for contexts 1, 3, 5 and 6, and a grade from 0 – 8 for contexts 2, 4 and 7.

Relative clause and relevant morphology rating scale

Section B: Grammaticization; Morphological Adequacy Scale

Descriptor	**Definition**	**Example**	**Points**
Target-like use of 3rd person singular & targetlike RC	Two instances of targetlike use of of 3rd person singular that compliments targetlike relativization	*'He thinks he likes the dog which has long ears'*	9
Target-like use of 3rd person singular & targetlike RC	One instance of targetlike use of 3rd person singular that compliments targetlike relativization	*'He wants the dog which has long ears'*	8
Target-like suppliance of RC only	Two instances of 3rd person singular that contain errors that compliment a targetlike	*'He thinks he like the dog which has long ears'* *'He thinks she likes the dogs*	7

	relative clause that contains no errors.	*which have long ears'* '*He thinks he like the dog which woman is looking at'*	
Target-like suppliance	A relative clause that exhibits targetlike relativization; contains no errors relating to verb tense but may contain other errors such as articles	'*He want the dog which has long ear'* '*He want the dogs which have long ear'*	6

Section A: Syntacticization; Relative Clause scoring scheme (Mochizuki & Ortega, 2008, p. 22)

Descriptor	**Definition**	**Example**	**Points**
Target-like suppliance	A relative clause that exhibits targetlike relativization; it may contain one or more errors that are irrelevant to the target structure, such as verb tense or the use of articles	*'I want the dog which have long ear'*	5
Developmental suppliance	A relative clause that contains any of four error types (i.e. pronoun retention, nonadjecency, incorrect relative marker, and inappropriate relative pronoun omission) described in the previous studies on relative clauses (e.g. Izumi, 2003)	*'I want the dog which many people are watching dog.'* *2. 'The dog is friendly which has long hair.'* *3. 'Ken likes the dog who has long ears.'* *4. 'I like the dog has long ears.'*	4
Attempt with processing overload	Relative clause attempted but containing a breakdown such as omission of head noun or verb in the relative	*'She wants which has long ears.'* *'She wants the dog which long ears.'*	3

	clause; these cases were frequent in the sample but have not been attested in previous studies of English relativization		
Least successful attempt	Relative clause where both developmental and processing load errors combine to cloud the success of the product and hinder intelligibility	*'Kanako wants to buy which has long hair and long ear dog.'*	2
Simplification	An utterance in which the participant tried to convey meaning without attempting relativization, alternative structures; these include either the structure derived from a direct translation form Japanese or alternative structures in English	*1.'long the dog that has long ear.'* *2. 'the dog with long hair.'*	1
Avoidance of Content	Formulation of the content involved in one of the seven contexts for obligatory suppliance was not attempted		0

Instructions for calculating accuracy:

1. Calculate each student's narration for each pre- and post-test. For example, start with the pre-test, enter the seven obligatory context scores for each student into excel, then sum the total for each student's pre-test score.
2. Align the students into their respective groups (GP and GUP) in SPSS.
3. Enter the raw value of each student's total pre-test score into SPSS.

4. Calculate the mean average score of each group. This is the group's pre-test accuracy score.

5. Repeat steps 1 to 4 to calculate each group's intermediate and delayed post-test accuracy scores.

Appendix S: Interator reliability results

	N 1	N 1	N 2	N 2	N 3	N 3	N 4	N 4	N 5	N 5	N 6	N 6	N 7	N 7	N 8	N 8
	Int. 1	Int. 2	Int. 1	Int. 2	Int. 1	Int. 2	Int. 1	Int. 2	Int. 1	Int. 2	Int. 1	Int. 2	Int. 1	Int. 2	Int. 1	Int. 2
Context 1	1	1	1	1	8	9	8	9	8	9	5	7	8	9	5	5
Context 2	1	1	5	4	7	6	8	8	6	6	5	4	7	8	8	8
Context 3	1	1	1	1	8	9	5	5	8	9	5	5	8	8	7	5
Context 4	1	1	1	1	7	7	5	5	8	8	7	8	8	8	7	6
Context 5	1	1	1	1	8	9	5	5	8	9	5	5	8	9	8	7
Context 6	1	1	1	1	6	5	1	1	7	6	8	6	1	6	6	4
Context 7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Appendix T: Instructions for coding and calculating syntactic complexity

Relative clauses per AS-unit:

Total number of relative clauses
Total number of AS-units

AS-units

An AS-unit = "an independent clause or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster, Tonkyn & Wigglesworth, 2000, p.365).
An independent clause is a clause that contains a finite verb.

Examples of an AS-unit (adapted from Gilabert, 2007a):

A man arrives at the house = 1 AS-unit
He is happy = 1 AS-unit
At the time she was waiting there comes a bus = 1 AS-unit
The car was driving fast but had no lights = 1 AS-unit
The dog which has long hair is happy = 1 AS-unit
The man = 0 AS-unit.

Subordinate clauses or sub-clausal units are associated with an independent clause provided the speaker has not paused more than 0.5 seconds, for example;
Paul went to the park (1.0 second) and he took the 14 bus = 2 AS-units.

Examples of sub-clausal units (Foster, Tonkyn & Wigglesworth (2000, p. 366):
A: How long you stay here
B: three months

Although these phrases contain grammatical errors, they have pragmatic meaning within the context and therefore each utterance would be considered an AS-unit.

Repeated clauses and incomplete clauses are not counted as AS-units, for example, the underlined structure;
'he likes the dog which has black hairs, which has black hair.' = 1 AS-unit

Relative clauses

Relative clauses are counted providing a relative pronoun is used, for example;
The dog which has long ears = 1 RC
The dog which long ears = 1 RC
The dog with long ears = 0 RC

Repeated relative clauses are not counted, for example,
The dog which has long ears which has long ears = 1 RC

Instructions to count AS-units using CLAN (adapted from Gilabert, 2007a):

1. Use the narrative file already coded for fluency speech rate A (see appendix U below), and remove the space bar syllable coding.
2. Enter 'ASU' at the end of each AS-unit.
3. Save the file.
4. In CLAN, click 'CTRL + D'
5. Select the file which has the narrative coded for AS-units.
6. Enter the formula: 'mlt + tSUB'
7. Click 'FILE IN' and select the file, then click 'RUN'
8. An 'Output' file will be generated and at the bottom a heading will appear as 'Number of Utterances' and the figure next to it will be the number of AS-units.

Instructions to count relative clauses using CLAN:

1. Use the narrative file already coded for AS-units above.
2. Remove the 'ASU' labels in the text and enter 'RC' next to each relative clause.
3. Save the file.
4. In CLAN, click 'CTRL + D'

5. Select the file which has the narrative coded for relative clauses.
6. Enter the formula: ‘freq +tSUB + sRC ’
7. Click ‘FILE IN’ and select the file, then click ‘RUN’
8. An ‘Output’ file will be generated and at the bottom a number will appear with RC next to it and that will be the number of relative clauses.

Relative clauses per AS-unit:

$$\frac{\text{Total number of relative clauses}}{\text{Total number of AS-units}}$$

Repeat the above steps to calculate each student’s pre- and post-test score for syntactic complexity. Then calculate the mean average for each group for each pre- and post-test.

Appendix U: Instructions for coding and calculating fluency (in line with Gilabert, 2007a)

Fluency speech rate A:

Total number of syllables x 60
Total number of seconds

Fluency speech rate B:

Total number of syllables (excluding false starts, repetitions, self-corrections, L1 use and incomprehensible language)
Total number of seconds x 60

Speech rate A: Examples of syllables for speech rate A (in line with Gilabert, 2007a):

1. The initial 's' is not a syllable, for example; 'strange' = one syllable.
2. Past tense 'ed'is not a syllable, for example, 'looked' = one syllable.
3. Present continuous, for example, 'looking' = two syllables.
4. 'doesn't' or 'didn't' = two syllables.
5. Japanese words are counted, for example 'etto the man' = four syllables
6. A false starts are counted, for example, 'the wo the man' = four syllables.
7. Repetitions are counted, for example, 'the the man' = three syllables.
8. Self-corrections are counted, for example, 'He think that she the man' = six syllables.

Speech rate B: Examples of syllables for speech rate B (in line with Gilabert, 2007a):

1. Remove the following underlined words that appear in self-corrections, repetitions and false starts for example;

Repetition: 'the the man is sad'
False start: 'the ca the dog'
Repair (self-correction): 'The woman err the man went to the shop'
The following is not counted as a repetition or self-correction:
'I think he the man is sad'

2. Remove syllables which appear in Japanese, for example;

‘Etto the man’ becomes ‘the man’

Instructions for speech rate A or B:

1. Enter the narration into CLAN.
2. If using speech rate A; divide all the words into syllables using the spacebar. If using speech rate B; divide all the words into syllables using the space bar excluding the examples outlined in speech rate B above.
3. Count the seconds using CLAN as follows:
4. First, link the audio file to the transcript by placing the audio file and the transcript in the same directory.
5. Open your CLAN transcript
6. Enter @medialine on the top left screen
7. Click Transcribe sound
8. Clan will ask for the audio file so click on the audio file
9. The audio should then start, when it does, click the spacebar after each utterance. This creates a ‘bullet’ at the end of each line which connects the audio to your transcription.
10. If you choose to stop at any point, to re-start, place your cursor on the last bullet and click ‘transcribe sound’ and the audio will start from the last recording of that bullet point, and whenever you press the spacebar, new bullet points will appear.
12. When your transcription is finished, click file, save and this will save the bullet points into your transcription.
13. To see the seconds, click Mode, expand bullets, then all the milliseconds for each utterance will appear at the end of each utterance where the ‘bullets’ were located.
14. Add up the milliseconds for each utterance and sum the total seconds of the transcription.
15. Counting syllables:
16. Click on ‘CTRL + D’ to open the command window.
17. Type in the formula ‘freq +tSUB’ then click ‘FILE IN’ and click the file of your narration.
18. An ‘Output’ file will then be generated and it will display all the syllables in the file. At the end of the file, there will be a heading “total number of words” and the figure next to it will be the total number of syllables.

19. Use the formula outlined at the top of the document to calculate each learner's pre- and post-test fluency score. Then calculate the mean average for each group for each pre- and post-test.

Appendix V: Main study interview transcriptions

Guided planner: 112

Guided and Unguided planner: 214

Week 2: 112

How was this task for you? How was it?

It was little easy.

A little easy? Ok, Why?

Because story is simple.

The story was simple. Ok, now when you were planning, how did you plan?

Umm, how?

Yeah, so when you were planning what was your focus?

Using this grammar.

Ok, can you give me an example?

For example, umm, using which, umm, which thinks ones

Ok, good, now when you were planning, did you think about grammar or vocabulary or the story?

Umm grammar

Grammar, just grammar?

And story

And story ok, when you were speaking, what was your focus when you were speaking?

Umm, to tell the story exactly

Tell the story exactly, ok, can you give me an example?

Umm, umm, boys are looking, girls are looking

Ok, good

Week 3: 112

How was this task? How was it for you?

It was little difficult more than last test

Ok, why?

Because, umm, because six, umm, six scene has two people.

Ok, I see. How did you plan what did you do when you were planning?

Using these girls

Ok, can you give me an example?

Using thinks which ……

Ok when you were planning your focus was?

I focused on explaining the, umm, umm, I focused on the explaining the car

Explain the cars, ok, when you were planning did you think about the vocabulary or the story or just grammar?

All of them

All of them, I see. Now, did you plan differently compare to last time?

No

The same?

Yes

Can you give me an example?

I used only thinks once and which

Ok, right, when you were speaking what was your focus when you were speaking?

I focused on using grammar, umm, exactly

Ok alright,

Week 4: 112

How was this task for you?

It was little difficult

Little difficult ok why?

Er because erm the car or something have two colours

Ok and when you were planning what was your focus? When you were planning?

I focused on using grammar

Grammar For example?

Using thinks amd likes and which was wants

Ok and when you were planning did you plan different compared to last week?

Erm almost the same

Almost the same, for example?

Erm I used the man is looking the girl is looking

Ok and when you were speaking what was your focus? When you were speaking

I focused on using grammar correctly

Grammar correctly, for example?

Erm for example think I I used thinks and wants

Ok

Week 2: 214

How was that speaking task, how was it for you?

Umm, It was difficult.

It was difficult? Why?

Umm, Umm, Umm, I can use word which.

Ok, now did you plan before you did the speaking task?

Before?

Yes, did you plan? Make notes? Was, it was easier?

It was easier

It was easier ok, right. And how did you plan what did you do when you were planning? When you planning, what were you thinking of? What did you do?

Fuu

When you were making notes

Yes,

Ok, were you thinking about grammar or vocabulary or the story, what were you thinking of when you were making notes?

Ahh, I wrote grammar.

You were thinking of grammar. ok just grammar anything else?

Ah, And story

Story, ok, why?

Umm Umm to tell story umm grammar is important.

Ok, when you were focusing on planning, your focus was grammar?

Oh yes,

Ok, thank you very much.

Week 3: 214

How was this for you? How was it?

Difficult

Difficult? Why?

Umm, these pictures, umm, I don't know.

Ok, alright. when you were planning what was your focus?

Focus, umm, story

Story, for example?

Umm, who is want

Ok and when you were planning did plan different compare to last time?

Ahh, I think similar

Similar? For example?

Umm, she think

Ok, when you were speaking what was your focus?

Umm, focus, umm, story

Story, ok

Week 4: 214

So how was this for you?

Umm Difficult

Difficult? Ok, why?

Umm these, these picture how describe, umm, I don't know

Ok, when you planning what was your focus?

Story, story

For example?

Umm, who wants what cars

Ok and when you were planning did you plan different compare to last time?

ahh, no

Same?

Almost

Almost. For example?

Umm, she she wants the cars is same

Ok right and ahh when you were speaking what was your focus?

Umm story and umm car's color and color car's color

Ok, I see good thank you

Appendix W: Most frequently used words of the GP and GUP group's post-task interviews (main study)

GP group

Week 2		Week 3		Week 4	
Frequency	**Word**	**Frequency**	**Word**	**Frequency**	**Word**
26	Grammar	16	which	14	which
12	Story	13	grammar	12	grammar
11	This	9	this	11	for
11	And	9	Is	10	same
10	Which	8	yes	8	use
9	example	8	use	8	think
7	To	8	no	8	and
6	vocabulary	7	has	7	to
6	Use	6	thinks	6	thinks
6	thinking	6	same	6	story
6	That	6	and	6	is
6	Or	5	focus	6	have
6	Easy	5	car	6	has
5	Wrote	4	To	6	example
5	Want	4	think	5	yes
5	Think	4	story	5	was
5	sentence	4	picture	5	wants
5	rabbits	4	On	5	this
5	How	3	vocabulary	5	so
5	Focus	3	using	5	or
4	Yes	3	So	5	on
4	Of	3	Or	5	etto
4	My	3	only	4	picture
4	Is	3	little	4	no
4	For	3	have	4	it
3	Yeah	3	for	4	i'm
3	Went	3	focused	4	er
3	understand	3	example	4	ah
3	thinks	3	difficult	3	you
3	They	3	different	3	word

3	So	3	bubble	3	want
3	sentences	2	when	3	using
3	picture	2	wheel	3	plural
3	One	2	wants	3	order
3	Long	2	want	3	looking
3	Has	2	unnto	3	focused
3	At	2	try	3	focus
3	about	2	three	3	car
2	writing	2	sentence	2	write
2	Using	2	recline	2	when
2	There	2	one	2	used

GUP group

Week 2		**Week 3**		**Week 4**	
Frequency	**Word**	**Frequency**	**Word**	**Frequency**	**Word**
33	grammar	14	grammar	16	Same
15	This	11	Is	14	And
13	Is	10	Which	11	Story
11	Story	10	Story	10	Car
11	Focus	9	To	7	On
10	To	7	Same	7	No
9	And	7	Focus	7	Is
8	Which	6	vocabulary	7	ichi
8	sentence	6	Last	7	grammar
7	Yes	6	example	6	wants
7	Think	6	Car	6	time
5	Long	5	Yes	6	for
5	How	5	On	6	example
4	Who	5	He	5	which
4	What	5	For	5	to
4	Want	5	And	5	this
4	vocabulary	4	Week	5	thinks
4	Use	4	Want	5	sisters
4	Thinks	4	Use	5	of

4	Just	4	This	5	cars
4	If	4	Think	4	so
4	First	4	different	4	or
4	At	4	colours	4	last
3	Wrote	3	Time	4	focus
3	We	3	thinks	4	difficult
3	Used	3	That	4	different
3	That	3	No	4	colour
3	Rabbit	3	Many	4	can
3	practice	3	character	3	what
3	picture	2	With	3	week
3	Only	2	What	3	use
3	On	2	wants	3	think
3	looking	2	remember	3	sarah
3	Likes	2	picture	3	do
3	imagine	2	Or	3	difference
3	Has	2	One	2	yes
3	Did	2	nothing	2	who
3	Before	2	Not	2	used
2	Yeah	2	nanka	2	today
2	Words	2	Make	2	they
2	Very	2	Like	2	there