

School of Education

Task Difficulty in Dialogic Oral Production by Indonesian EFL Learners

Mahpul

**This Thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University**

April 2014

Declaration

To the best of my knowledge this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signed: _____

Date: _____

Acknowledgements

I am very grateful to the individuals and institutions for assistance with this thesis. First, I would like to sincerely thank my principal supervisor, Professor Rhonda Oliver for her eminently clear guidance, constructive feedback, and ‘massaging’ the thesis in the process of writing. I would have been lost without her guidance. My sincere gratitude also goes to Dr. Christopher Conlan for his guidance and comments in the early process of writing the thesis. I also thank my co-supervisor, Associate Professor Jenny Jay for her editing. My sincere thanks are extended to Dr. Anna Alderson for her time and help concerning my language use.

Many thanks also go to my scholarship sponsor, DIKTI and my home University, the University of Lampung for their support of my study. I also thank the students of the English Study Program, the University of Lampung commencing in 2010 for their participation in this study. My thanks are extended to Gilang Santoso and Mei Dianita for their help converting the recording data into a form I could use and Rini Fadillah Situmorang who assisted me verify final transcripts of the data. I also thank L1 speakers of Australian English (who cannot be named for confidential reasons) for their participation in this study.

Special thanks go to my wife, Esti Winarti and my three beloved sons, Vergie Winanda, Pradipta Priandaru, and Inevito Adyatma for their care, love, support, and patience during my absence. Last, but not the least, my thanks and respect to my late parents, Father Dulrohmat and Mother Samijem, I do thank you both for your love, support, and struggle with me. Rest In Peace my beloved parents.

Preface

My son pulled out from an English course due mainly to boredom with the teaching-learning process in the classroom. He gave reasons indicating that he found it boring to memorize lists of vocabulary and have “discrete-grammar teaching” without being able to use English in the context of real communication. This situation is exacerbated by the fact that learners’ language performance has been primarily assessed by measures, which are linguistic rather than communicative such as the appropriateness of utterances within a specific context.

Drawing on this I was curious to understand whether providing learners with a task-based language teaching, which primarily focuses on meaning rather on forms, enables learners to engage in genuine communication. I am also interested in exploring the assessment of learners’ language performance, particularly in relation to complexity, accuracy and fluency (CAF) as discrete measures, which are traditionally used to measure learners’ language output and L1 raters’ judgments which might view learners’ language performance holistically.

Abstract

The increasing use of task-based approaches in language teaching has resulted in a number of studies exploring learners' language performance according to various task conditions, including learner performance of monologic (one way flow of communication) and dialogic (two way flow of communication) tasks. Task studies have been conducted from different perspectives, one of which is the information-processing perspective. This perspective is primarily to do with the manipulation of task complexity (cognitive factors). This current study was informed by Robinson's Cognition Hypothesis.

Pedagogically task complexity is seen to be the main basis for task design. Task complexity involves the manipulation of a learner's cognition along two dimensions: the resource-directing; and the resource-depleting dimensions. That is, task can be made simpler, requiring less cognitive engagement, or more complex, requiring greater cognitive engagement. For example, according to Robinson's Cognition Hypothesis tasks can be manipulated to include more or less planning time, a greater or fewer numbers of elements in the task, and so on and done so in a way that either increases or decreases the level of difficulty.

To measure the effect of task complexity investigations have examined learners' language production, mostly in terms of complexity, accuracy, and fluency (CAF). These were the measures also used in the current study. Complexity was calculated in terms of Syntactic Complexity, Percentage of Lexical Words, and Guiraud's Index of Lexical Richness. Accuracy was assessed using Error-Free AS-Units, Percentage of Self-Repairs, and Ratio of Repaired Errors to Unrepaired Errors. Fluency was measured in terms of Unpruned Speech Rate A and Pruned Speech Rate B.

However, a number of limitations with regard to task-based CAF studies have been identified, including that: i) most studies have focused primarily on monologic rather than on dialogic tasks, and on either the resource-directing or the resource-depleting factors; ii) very few studies have investigated participants' in-depth perceptions of task

complexity; and, iii) the use of CAF to measure learners' language production has recently been subject to debate. Therefore, tasks in this study were made more or less complex by simultaneously manipulating the provision (or not) of planning time and decreasing/increasing the number of elements the learners should consider (i.e., the resource-dispersing and the resource-directing dimensions). Further, unlike a number of previous studies, the tasks in the current research were dialogic in nature – that is they were interactive tasks requiring a two way exchange of information. This study also explored the relationship between participants' perceptions of task difficulties (affective factors) and task complexity. Finally, the degree of fit between the Indonesian participants' oral production as measured by CAF and L1 raters' judgments of oral production was measured.

The findings of this research contribute to our understanding of task complexity, especially from an information-processing perspective, which in turn can be used to inform the implementation of task-based approaches. Theoretically the results suggest that the manipulation of task difficulty (i.e., cognitive factors) along resource-dispersing and the resource-directing dimensions only partially supported the predictions of the Cognition Hypothesis. Learner performance was also influenced by interactive, learner, and input factors. This was especially reflected in the learners' perceptions of the tasks. Together these quantitative and qualitative findings highlight the dynamic relationship between tasks, their complexity and learners' performance.

In terms of L1 raters' judgment of Indonesian EFL oral production and CAF measures, there is little evidence to suggest a strong degree of fit. This finding contributes to body of evidence highlighting the complexity surrounding measures of learners' performance. It also draws into question the appropriateness of using CAF measures alone to determine learner performance.

There were a number of limitations in the current research. In particular, the repetition effect of similar tasks and the limited number of participants seemed to have impacted on the results. This will need to be addressed in future research.

However, the results of the current study can be used to inform the development of a framework for designing pedagogical tasks. This framework will support EFL teachers and syllabus designers to design appropriate tasks according to learners' need.

Table of Contents

Declaration	ii
Acknowledgements	iii
Preface	iv
Abstract	v
Table of Contents	viii
Lists of Figures.....	xi
Lists of Tables	xii
Chapter One INTRODUCTION	1
1.1 Teaching of English in Indonesia	1
1.2 Task-Based Language Teaching.....	4
1.3 Research Objectives.....	8
Chapter Two REVIEW OF LITERATURE	9
2.1 Overview.....	9
2.2 Task-Based Approach in Language Teaching	9
2.3 Definition of Tasks in Language Teaching	11
2.3.1 Real World Tasks	12
2.3.2 Pedagogical Tasks	13
2.3.3 Dimensions of Tasks.....	15
2.4 Task-Based Research.....	18
2.4.1 Tasks and information-processing	20
2.4.2 Skehan's Limited Capacity Model/Cognitive Approach.....	23
2.4.3 Robinson's Cognition Hypothesis	24
2.5 Perceptions of Task Difficulty	38
2.6 Measuring Complexity, Accuracy, and Fluency (CAF).....	39
2.6.1 Complexity	40
2.6.2 Accuracy	43
2.6.3 Fluency	45
2.7 Holistic Assessment of Language Performance: L1 raters' judgments.....	48
2.8 Summary.....	51

Chapter Three METHODOLOGY	53
3.1 Overview.....	53
3.2 Design of the Research	53
3.3 Data Collection Procedures	55
3.4 Participants	58
3.5 Materials and Equipment.....	59
3.5.1 Tasks.....	59
3.5.2 Interview Protocols.....	63
3.5.3 The L1 Rating Scale	63
3.5.4 Trialing the Instruments.....	64
3.6 Data Analysis.....	68
3.6.1 Analysis of Participants' Oral Production about Tasks.....	68
3.6.2 Analysis of participants' post-task interviews.....	72
3.6.3 Analysis of L1 raters' Assessments.....	73
3.7 Summary.....	74
Chapter Four FINDINGS ABOUT TASK DIFFICULTY	75
4.1 Overview.....	75
4.2 CAF and the Four Levels of Task Difficulty.....	75
4.3 Comparing Planned and Unplanned Tasks.....	82
4.4 Comparing Simple and Complex Tasks	89
4.5 Comparing Planned Simple and Unplanned Complex Tasks and Unplanned Simple and Planned Complex Tasks	95
4.6 Summary and conclusion: Comparisons of task conditions according to CAF measures	100
Chapter Five FINDINGS OF INDONESIAN PARTICIPANTS' PERCEPTIONS OF FOUR LEVELS OF TASKS	102
5.1 Overview.....	102
5.2 The Indonesian Participants' Perceptions of Task Complexity.....	102
5.2.1 Task Difficulty.....	104
5.2.2 Degree of Stress	112
5.2.3 Degree of Confidence	117
5.2.4 Participants' Interest	122
5.2.5 Participants' Motivation	126
5.2.6 Learning opportunities.....	130

5.2.7 Dialogic Nature of the Tasks	133
5.3 Summary.....	135
Chapter Six FINDINGS: L1 RATERS' JUDGEMENTS ON INDONESIAN PARTICIPANTS' PERFORMANCE	136
6.1 Overview.....	136
6.2 L1 Australian English Raters' Assessments of the Oral Production of the Indonesian Participants.....	136
6.2.1 Comparisons of L1 Raters' Judgments and Complexity Measures.....	143
6.2.2 Comparisons of L1 Raters' Judgments and Accuracy Measures	145
6.2.3 Comparisons of L1 Raters' Judgments and Fluency Measures.....	148
Chapter Seven CONCLUSIONS	151
REFERENCES.....	157
APPENDICES	170
Appendix 1	171
Appendix 2.....	172
Appendix 3.....	173
Appendix 4.....	174
Appendix 5.....	175
Appendix 6.....	176
Appendix 7	177
Appendix 8.....	178
Appendix 9.....	179
Appendix 10.....	180
Appendix 11	181

Lists of Figures

Figure 1: Components of Task (from Nunan, 1989, 2004).....	15
Figure 2: Robinson’s Triadic Componential Framework	25
Figure 3: Ellis’ (2005) model of task-based planning.....	30
Figure 4: Diagram of preparation stage	56
Figure 5: Three phases of the implementation stage.....	57
Figure 6: Four levels of (dialogic) task difficulty developed in this study	60
Figure 7: Means of Syntactic Complexity for the four levels of tasks.....	77
Figure 8: Means of Percentage of Lexical Words for the four levels of tasks.....	77
Figure 9: Means of Guiraud’s Index generated by four levels of tasks.	78
Figure 10: Means of Accuracy, Error-Free AS-Units, for the four levels of tasks.	79
Figure 11: Means of Accuracy, Ratio of Self-Repairs to Unrepairs and Percentage of Self-Repairs for the four levels of tasks.	79
Figure 12: Means of Fluency Speech Rate A and Speech Rate B for the four levels of tasks.	80

Lists of Tables

Table 1: Configurations of the binary code (+/-) of task difficulty (Cognitive factors)	28
Table 2: Calculation of Lexical Words	69
Table 3: Measures of Accuracy.....	69
Table 4: Measures of Fluency	71
Table 5: Comparisons of four levels of task difficulty	71
Table 6: Means of CAF for the four levels of Tasks.....	76
Table 7: Repeated measures ANOVA: main effects of different levels of tasks for CAF measures.....	81
Table 8: Mean differences of planned and unplanned tasks for Complexity Measures ..	82
Table 9: Mean differences of planned and unplanned tasks for three Accuracy Measures	86
Table 10: Mean differences of planned and unplanned tasks for two Fluency Measures	88
Table 11: Mean differences of simple and complex tasks for Complexity measures.....	89
Table 12: Mean differences of simple and complex tasks for three Accuracy measures	92
Table 13: Mean differences of simple and complex tasks for two Fluency measures.....	93
Table 14: Mean differences between Tasks 1 and 4, and Tasks 2 and 3 for Complexity measures.....	95
Table 15: Mean differences between Tasks 1 and 4, and Tasks 2 and 3 for Accuracy measures.....	97
Table 16: Mean differences between Tasks 1 and 4, and Task 2 and 3 for Fluency	99
Table 17: Summary of the effect of planning time and the number of elements.....	101
Table 18: Participants' agreement and disagreement about the complexity of four tasks	103
Table 19: Participants' perceptions about the of difficulty of four levels of tasks	105
Table 20: Participants' perceptions about the stress tasks	113
Table 21: Students' perceptions about their confidence with tasks	118
Table 22: Participants' interest in the four levels of	

tasks.....	123
Table 23: Students' perceptions about tasks and their motivation.....	127
Table 24: Participants' perceptions about tasks and their learning opportunities.....	131
Table 25: Participants' perceptions about dialogic features of the four tasks.....	133
Table 26: Three L1 raters on Likert scales for the four levels of tasks.....	136
Table 27: Rating agreements of three L1 raters on oral production for four levels of tasks	137
Table 28: L1 raters' written comments on the Indonesian oral production for four versions of tasks.....	138
Table 29: Comparisons of L1 ratings and Complexity measures for four versions of tasks	144
Table 30: Comparisons of L1 ratings and Accuracy measures for four versions of tasks	146
Table 31: Comparisons of L1 ratings and Fluency Rates A and B for four levels of tasks	148

Chapter One

INTRODUCTION

1.1 Teaching of English in Indonesia

English has long been recognized by the majority of Indonesian people, not only as a *lingua franca* among communities worldwide, but also as the language of commerce, science and technology. The status of English, as Lauder (2008) notes, has mainly been about serving the country's development needs. Consequently, for some time English has been acknowledged as the first foreign language that must be formally taught to students from junior high school up to university. In addition, English is also introduced as an elective subject to students at elementary school. Thus, English teaching in Indonesia has been generally taught to students within all three levels of education: in elementary schools as an elective subject, in high schools from year seven to year twelve as a compulsory subject, and, at university level as a compulsory subject particularly for the first year students.

As indicated, at the elementary school level teaching of English is optional. It is introduced to students from year four to year six. However, even at this level the teaching-learning activities commonly pay more attention to engaging learners in memorizing vocabulary lists and grammatical rules, rather than providing them with meaningful activities that enable them to communicate in English. This means that the teaching of English at the elementary school level has been regarded by some as being unsatisfactory due to the use of inappropriate language teaching methods. The reason this occurs is because the majority of English teachers teaching at elementary schools neither have backgrounds of teaching English to speakers of other languages (TESOL) nor can they communicate well in English themselves (Kasihani, 2010; Zein, 2010). Poorly qualified teachers using inappropriate language teaching methodologies and teachers' low English proficiency, therefore, have been identified as the major issues for English teaching at the elementary school level in Indonesia.

Unlike English teaching at the elementary school level, the teaching of English to students at the high school level is obligatory. It aims at ensuring the students have the capability of communicating in English in both spoken and written forms. To achieve this goal, a series of up-grading programs, workshops, seminars, and in-service training for English teachers have been put in place in an attempt to find appropriate models of language teaching. However, the teaching practice in the classroom setting has remained largely 'traditional': that is, language forms and reading skills (receptive skills) have been prioritized, while oral skills (speaking) have been given the least attention or even neglected altogether.

This might be because the teachers must prepare their students for national examinations which, in the main, measure accuracy (language forms) and reading comprehension, rather than speaking skills. In fact, teachers are regarded as successful in their jobs if the majority of their students pass the examinations. Thus, they are most likely to use methods that emphasize "discrete-grammar teaching" to their students and provide them with a number of grammar exercises rather than engaging them in learning activities that allow for genuine communication. Consequently, while the students may "know" the language forms, they cannot effectively communicate in English even at a very basic level (Setiyadi, 2009; Kasihani, 2010). This is in spite of the fact that they have formally learned English at junior and senior high schools for six years.

The teaching of English at the university level can be classified as being of two types: i) English for students whose major is other than English, such as mathematics, law, agriculture, etc., and, ii) English for those whose major is English. The former refers to the teaching of English for specific purposes (ESP) in that the emphasis is given to developing students' vocabulary and understanding of the texts related to their field of study. This is intended to enable them to access publications in English to improve their background and content knowledge. These students are obliged to take English classes for one or two semesters. The latter group of students are in English programs primarily aimed at preparing them to be English language teachers. These students study language skills (listening, speaking, reading, and writing), linguistic theory (syntax, pronunciation and vocabulary) especially in the first two years, and other subjects or

units related to the principles of teaching English as a foreign or a second language (TEFL/TESOL). As such, these students are required to demonstrate not only competence of communication in English in both spoken and written forms, but also in language teaching methodology. However, there is evidence that the teaching of English to both groups of students has been unsuccessful in the sense that the majority of the university graduates have been reported as still being unable to speak English well (Setiyadi, 2009; Saragih, 2009). Therefore, overall there is unsatisfactory achievement at the elementary, high school and university levels in Indonesia.

The low levels of ability in spoken English appears to have become a major issue, not only among the Indonesian learners of English, but also among other Asian students particularly those coming from countries where English is a foreign language. For instance, studies have revealed that the anxiety of speaking English and lack of speaking practice both inside and outside the classrooms are seen to have contributed to the low English proficiency levels among Asian students (Tsai, 2003; Rahim, Ahmad, & Rosly, 2004; Na, 2007). This suggests that these learners might not have been provided with the type of learning activities that facilitate their development in speaking English. Accordingly, the students need to be provided with learning activities that stimulate their speaking. Samuda and Bygate, (2008, p. 7) argue that one way of engaging students language use is through task-based approaches.

However, as Luciana (2005) points out, task-based language teaching approaches have not been widely adopted in Indonesia. Luciana suggests that English teachers are reluctant to adopt task-based language teaching (TBLT) firstly because the majority doubt that it is an approach that enables students to improve their language development, and, secondly, because the teachers lack confidence in terms of the implementation of TBLT in their classrooms. These problems might be because Indonesian English teachers insist on explaining linguistic rules and providing students with reading texts rather than engaging them in speaking activities. Generally a lack of experience with TBLT appears to be a major constraint for shifting teachers from their ‘traditional’ practice, which “focuses on forms,” (also known as “discrete-point grammar teaching”,

Long, 2000, p. 179) to TBLT which puts emphasis primarily on meaning rather than on forms (although it must be noted that a “focus of form” is also pivotal to TBLT).

In summary, despite a long history of teaching English in Indonesia, English teaching is regarded as unsuccessful at almost all of the education levels from the elementary school to university. The low level of speaking ability of students has become a major issue in Indonesia. Explanations for this lack of success point to the classroom practices of the teachers. Clearly there is a need to shift the paradigm of English teaching in Indonesia - from predominantly form-focused to meaning-focused activities, and this might be able to be achieved through TBLT.

1.2 Task-Based Language Teaching

In the last few decades, task-based language teaching (TBLT) has been widely and extensively adopted as an approach to language teaching, for example, in China, Japan and India (Prabhu, 1987; Nunan, 2004; Robinson, 2003, 2005, 2007). TBLT focuses on using language as a means of communication in contrast to the traditional practice of language teaching which insists primarily on form-focused learning. Thus, the use of tasks in the classroom context is intended to provide learners with learning activities that reflect real-life situations in which language is naturally used as a means of communication.

The use of TBLT has recently resulted in a growing number of task-based studies (e.g., Skehan, 1996; Yuan & Ellis, 2003, 2005; Gilabert, 2005; Robinson, 2001, 2003, 2005, 2007a; Gass, Mackey, & Ross-Feldman, 2011; Oliver, Grote, Rochecouste, & Exell, 2013).

This field of research lies within the area of second language acquisition (SLA), and particularly the interactionist paradigm, and is based on theories which suggest the naturalistic exposure to and use of language is a prerequisite for language development (Skehan, 2003). Task-based studies have been approached from two different perspectives: the interactionist and the information-processing perspectives. The former

is based on the interaction hypothesis (Long, 1983, 1985, 1996). The latter is called the cognitive approach (Skehan, 1998) or the cognition hypothesis (Robinson, 2001a, 2001b, 2003, 2005, 2007a, 2007b).

Studies undertaken using an interactionist perspective have been mainly concerned with aspects of interaction, such as negotiation of meaning and more recently a focus on form and feedback (see Long, 1996). This is because interaction of various kinds is argued to be facilitative of language acquisition (again see Long, 1996, but also Mackey & Oliver, 2002; Oliver & Mackey, 2003; Mackey, Oliver, & Leeman, 2003; Mackey, Kanganas, & Oliver, 2007). However, the studies within this approach have been criticized for a number of reasons (Foster, 1998; Skehan, 2003), such as the data being based on laboratory-based studies rather than being undertaken in real classrooms.

Studies from the information-processing perspective have been mainly concerned with the investigation of tasks and cognitive factors and the differential effects these have on language performance, especially in terms of complexity, accuracy, and fluency (CAF). Extensive studies of the manipulation of tasks and task factors and the impact on CAF have been conducted in a number of different ways. For example, studies have investigated the effect of tasks while manipulating planning time (e.g., Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Yuan and Ellis, 2003, 2005; Philp, Oliver, & Mackey, 2006; Foster & Tavakoli, 2009; Ahmadian & Tavakoli, 2010); number of elements (e.g., Robinson, 2001a; Michel, Kuiken, & Vedder, 2007; Kim, 2009); number of elements and reasoning demand (e.g., Kuiken & Vedder, 2007, 2008; Gilabert, Baron, & Llanes, 2009); planning time and present and past references (e.g., Gilabert, 2005).

However, to date studies investigating tasks have been predominantly concerned with monologic (one-way) rather than dialogic (two-way) tasks (see Robinson, 2001a, 2001b, 2003, 2005, 2007a; Yuan & Ellis, 2003; Gilabert, 2005; Foster & Tavakoli, 2009; Ahmadian & Tavakoli, 2010; Tavakoli & Foster, 2011; Kormos & Trebits, 2012). Since the information performed through monologic tasks flows in one direction, no balance in roles between learners is expected to occur in the conversation. Consequently, a learner performing monologic tasks (the ‘speaker’) may dominate the other learner

(the ‘listener’) as the role of the listener in the dyad is merely to be a trigger, that is, asking questions rather than sharing ideas. Consequently, learners, and the listeners in particular, are expected to be less actively involved in conversations as the tasks are less interactive. On the other hand, the use of dialogic (two way) tasks has been found to stimulate learners to participate actively in conversation in a dyad and ultimately lead learners to improve their spoken language (Riggenbach, 1989; Anton, 1999; Bell, 2003; Michel et al., 2007). For this reason in the current study dialogic tasks were used.

Further, almost all of the studies regarding the manipulation of task factors undertaken within the information-processing perspective have been conducted along one dimension (i.e., the resource-directing or the resource-depleting dimension) rather than simultaneously investigating two dimensions (e.g., Robinson, 1995, 2001a, 2007a; Yuan & Ellis, 2003, 2005; Kuiken & Vedder, 2007, 2008; Ishikawa, 2007; Foster & Tavakoli, 2009; Ahmadian & Tavakoli, 2010). According to Robinson (2001a, 2001b, 2003, 2005, 2007a, 2007b), the resource-directing dimension refers to cognitive demands requiring attention and working memory that directs learners to focus on linguistic forms (e.g., a number of elements) whereas in the resource-depleting dimension it focuses on “performative or procedural demands” that require a learner’s attentional and memory resources, but the dimension does not direct learners to any particular linguistic form (e.g., planning time). A notable exception is the study undertaken by Gilabert (2005) who simultaneously investigated the roles of planning time and present and past activities (here and now), although, once again this was done with narrative (monologic) tasks. Michel et al. (2007) did compare tasks factors within monologic and dialogic conditions, however, only one dimension (the resource-directing) was investigated. Therefore, although numerous task-based studies have been conducted with a number of different dimensions and degrees of complexity, few, if any, empirical studies have been undertaken simultaneously along two dimensions (the resource-directing and the resource-depleting) within the dialogic tasks. The current study did this.

In addition, there appears to be a need to extend the lines of research with respect to learners’ perceptions of task difficulty. Robinson (2001a) did explore learners’ perceptions of task conditions, but did so using closed-ended questions on a nine-point

rating scale, and this quantitative approach has since been used in other studies (e.g., Gilabert, 2005; Kim, 2009). Since Robinson's model of questions was quantitative in nature, it might not have allowed researchers to explore learners' perceptions in depth. There is clearly a need for qualitative or open-ended questions in addition to the quantitative results to allow for the in-depth exploration of learners' perceptions (Robinson, 2001, 2003, 2005; Kuiken & Vedder, 2007; Foster & Tavakoli, 2009; Ahmadian & Tavakoli, 2010). This is the approach that was used in the current study.

Lastly, in a large number of studies CAF have been used as discrete measures of learners' language performance (Skehan, 1998; Robinson, 2001a, 2001b; Yuan & Ellis, 2003, 2005). However, the use of CAF to assess learners' language performance has recently been challenged in SLA research (Housen & Kuiken, 2009; Larsen-Freeman, 2009; Skehan, 2009; Norris & Ortega, 2009; Pallotti, 2009). Larsen-Freeman (2009), for instance, suggests that because of "the lack of suitable measures in SLA research" (p. 580). CAF are the default measures and this is problematic from the perspective which sees language production as a holistic phenomenon. Unfortunately few alternatives for how to assess learners' language performance on tasks have been proffered. One option suggested by Davies (2003, 2011) is the use of L1 speakers' judgments. However, there is a dearth of research about the utility of this within the TBLT area. Thus there is a need to explore the degree of fit between CAF and L1 speakers' judgments on learners' language performance and this in the final area that was addressed in the current study.

In summary, studies investigating the effect of task conditions on CAF have mainly been conducted using monologic (narrative) rather than dialogic tasks. In addition, task studies have predominantly investigated one dimension, namely the resource-directing (e.g., a number of elements) or the resource-dispersing (e.g., planning time), rather than simultaneously examining two dimensions. Learners' perceptions of task, in the main have been conducted by using quantitative approaches rather than exploring learners' in-depth perceptions. Finally, in most studies language performance has been determined using CAF measures, rather than L1 speaker judgments. On this basis this research sought to achieve the following Research Objectives.

1.3 Research Objectives

This study aimed to investigate dialogic (that is, two-way, interactive) task difficulty manipulated simultaneously within planning time (+/- planning time) and the number of elements (+/- few elements). This was undertaken with Indonesian learners of English. In addition, the participants' perceptions of task difficulty were explored in-depth. Finally, this study sought to map the findings of CAF onto L1 speakers' intuitions concerning performance in the target language. Therefore, this study specifically addressed three major inter-related issues:

- i) the effects of dialogic task complexity manipulated simultaneously along the two dimensions (i.e., the resource-directing and resource-depleting) of oral production of the Indonesian participants as measured by CAF;
- ii) the relationship between the participants' perceptions of task difficulty (learner factors) and the four levels of dialogic task difficulty (i.e., manipulated simultaneously with planning time and a number of elements); and
- iii) the degree of fit between CAF and L1 speakers' judgments.

The chapter that follows examines the literature in these three interrelated areas.

Chapter Two

REVIEW OF LITERATURE

2.1 Overview

In this chapter a review of the literature is provided. A specific focus of this is task-based approaches relevant to the current study. Therefore, this chapter first begins with a description of tasks and task-based teaching approaches are described. Next task-based research undertaken from the information-processing perspective is described. In particular, the Cognitive Approach, the Cognition Hypothesis and CAF measures of learners' language performance are outlined in detail. As native (L1) speakers' judgments were also used to measure learner performance in the current study, literature on this topic will also be discussed in this chapter.

2.2 Task-Based Approach in Language Teaching

Task-based approaches to language teaching, known as task-based language teaching (TBLT)¹ and also called task-based language learning (TBLL) or task-based language instruction (TBLI) began in the 1980s. They emerged from the constructs of communicative approaches to language teaching first crystallized by Brumfit and Johnson (1979) and subsequently developed by a number of others (e.g., Long, 1985; Long and Crookes, 1992; Crookes, 1986; Crookes & Long, 1987; Prabhu, 1987; Nunan, 1989, 2004; Ellis, 2003, 2005; Van den Branden, 2006; Van den Branden, Bygate, & Norris, 2009; Shehadeh & Coombe, 2010). As a consequence, the term *task* is now used to replace “communicative activities as the basic units of the communicative approach” (Skehan, 2003, p.1). This change was based on the belief that tasks are “one kind of holistic activity which can play a significant role in second language learning, teaching and testing” (Samuda & Bygate, 2008, p.7).

¹ Therefore the terms task-based approaches and TBLT are used interchangeably throughout this thesis

Within the TBLT “language pedagogy” paradigm, “language is enmeshed in human activity, rather than being a discrete and separate object of analysis” (Mickan, 2004, p.181).

Van den Braden et al. (2009) in their comprehensive review of TBLT contrast it with traditional, synthetic language teaching practices and make the point that TBLT is “a model of second language learning conceptualized in terms of holistic activities, meaning-based approaches, and learner-driven activities” (p.5). That is, in TBLT, learners are engaged in using the language for functional purposes by integrating different linguistic sub-skills (Van den Braden, et al., 2009). Further, meaningful language activities are the primary focus and learners are actively involved in opportunities to practice the language with other learners for functional purposes and the focus is not on the forms of language, but rather on making meaning. As such they enable second or foreign language learners to interact with other learners for genuine communication that may ultimately lead to the maximum use of the target language (i.e., the language the learners seek to acquire). Therefore, the shift from ‘traditional’ teaching practice to task-based learning is based on the belief that task-based approaches promote more effective language learning (Long, 1985; Swan, 2005; Shehadeh & Coombe, 2010). In task-based approaches learners are exposed to the activities of “a natural context for language use” (Larsen-Freeman, 2000, p. 144) which are argued to facilitate second language acquisition (SLA) (Larsen-Freeman, 2000; Richards & Rodgers, 2001; Beaven, 2005).

In response to an early call for the use of task-based approaches in language teaching, Prabhu (1987) developed a project known as the Communicational Teaching Project in Bangalore, India. This project was informed by SLA theory and based on the premise that language teaching practice which focuses primarily on language forms is not effective for developing learners’ competence in the target language. Instead he chose task-based approaches believing them to provide learners with appropriate activities which engaged them in language use. Specifically, Prabhu (1987) employed three types of tasks for the learners to perform: information gap tasks, reasoning tasks, and opinion tasks as a manifestation of the so-called “meaning-focused activities” (p.46). These latter tasks were found to promote more language learning than the other two types of tasks.

Based on this work Prabhu (1987) and then others, Larsen-Freeman (2000) described how task-based approaches provide learners with opportunities for interactions that enable learners “to work to understand each other, and express their own meaning, and listen to language which may be beyond their present ability” (p. 144). Further, she suggests that through task-based lessons learners are exposed to a wide variety of linguistic forms, rather than exposing them to a specific function or a particular form of language, such as occurs in traditional synthetic approaches.

In short, the development of TBLT has involved a paradigm shift in language teaching and learning from the traditional, synthetic approaches to language teaching which have a primary focus on “forms, discrete-learning, and teacher-centered activities” to task-based approaches which actualize language as a means of communication, one which “places the communication as the heart of teaching procedures” (Van de Branden et al., 2009, p. 5). This is because it is believed that task-based approaches in a classroom setting lead to successful second language (L2) learning.

2.3 Definition of Tasks in Language Teaching

‘Tasks’, in the context of language learning, have been defined in a number of ways by different researchers (Pica, Kanagy, & Falodun, 1993; Nunan, 1989, 2004; Van den Branden, 2006; Long & Norris, 2009). Drawing on a number of different definitions of tasks, Pica et al. (1993) characterized tasks in two ways: tasks oriented toward goals and tasks as work or activities. The former are intended for learners to achieve an outcome and to carry out a task with a sense of what they need to accomplish through their talk or action. The latter concerns learners’ active roles in performing tasks, whether they are working individually or in pairs or groups. Nunan (2004) and Long and Norris (2009) make two other basic distinctions: real-world or target tasks and pedagogical tasks. Target tasks refer to uses of language in the real-world beyond the classroom, where as pedagogical tasks are those that occur in the classroom. Similarly, Van den Branden (2006, p. 4) classifies tasks in terms of language learning goals and educational activities.

Therefore, while various authors may use slightly different terminology, what is important in terms of practice is that “real-world tasks” or “target tasks” can be transformed into “pedagogical tasks” to create learning opportunities in the classroom (Nunan, 2004, p. 19). In turn these tasks constitute a ‘bridge’ enabling learners to perform real-world tasks. The distinction between real-world tasks and pedagogical tasks is discussed in more detail in section 2.3.1 and 2.3.2 below.

2.3.1 Real World Tasks

Real-world or target tasks have been defined in various ways by different researchers. Long (1985) for example, defines a task in a general sense as:

A piece of work undertaken for oneself or for others, freely or for some reward. Thus, examples of a task include painting a fence, making an airline reservation, borrowing a library book, taking a driving test, typing a letter, weighing a patient, sorting letters, taking a hotel reservation, writing a cheque, finding a street destination and helping someone across a road. In other words, by ‘task’ is meant a hundred and one things people do in everyday life, at work, at play, and in between. Tasks are the things people will tell you they do if you ask them, and they are not applied linguists (p. 19).

Hence the definition of a *task* by Long does not necessarily correspond to language-learning activities. However, the definition does suggest that a *task* is intended for learners, as Ellis (2003, p.3) notes, “to achieve situational authenticity” as reflected in real life activities, which require either the use of language (e.g., writing a cheque) or without the use of language (e.g., painting a fence).

This was elaborated further in the definition provided by Samuda and Bygate (2008):

Holistic activities which engage language use in order to achieve some non-linguistic outcomes while meeting linguistic challenges, with the overall aim of promoting language learning, through process or product or both (p. 69).

Somewhat surprisingly these definitions of a task do not always focus on the achievement of linguistic outcomes. However, Van den Branden (2006, p. 4) does make the link between tasks and language use. Specifically he defines a task as “an

activity in which a person engages in order to attain an objective, and which necessitates the use of language”. What is clear from these various definitions is that the ultimate goal of TBLT is developing the capability of language learners to use the target language in order to perform real world tasks.

2.3.2 Pedagogical Tasks

In addition to real world or target tasks, are pedagogical tasks. Again these are defined in different ways by various researchers. Early on Nunan (1989), for instance, defined a pedagogical task as:

A piece of classroom work that involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form (p.10).

Thus, Nunan indicates that the focus of a task is primarily on meaning. Similarly, Skehan (1996) suggests a meaning-focused orientation when he defines a task as:

An activity in which meaning is primary; there is some sort of relationship to the real world; task completion has some priority; and assessment of task performance is in terms of task outcome (p. 38).

Similarly, Ellis (2003) defines a task as:

A work plan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed. To this end, it requires them to give primary attention to meaning and to make use of their own linguistic resources, although the design of task may predispose them to choose particular forms. A task is intended to result in language use that bears a resemblance, direct or indirect to the way language is used in the real world. Like other language activities, a task can engage productive or receptive, and oral or written skills and also various cognitive processes (p.16).

Therefore, the focus of tasks within TBLT is on meaning rather than on the forms of language (Long, 1991; Long & Robinson, 1998; Skehan, 1996; Ellis, 2003, 2005, 2009a; Nunan, 2004). (As indicated in Chapter One, p. 3, language forms refer to “discrete-point grammar teaching”, Long, 2000, p. 179).

According to Ellis (2005), the focus on meaning includes both pragmatic and semantic meaning. Ellis (2005) argues that pragmatic meaning which is “the highly contextualized meanings that arise in acts of communication” (p. 211) is more crucial to language learning than semantic meaning. This is because communicative competence, the goal of language learning, requires speakers to interact in ways that is both meaningful and appropriate for the context.

While the focus of TBLT is primarily on meaning and “learner-driven activity” in the sense that learners are actively engaged in interaction with other learners, the role of form is also recognized as important in language learning in general and in TBLT in particular (Van den Branden et al., 2009, p. 6). This is because meaning and form cannot exist separately as “the existence of form is to enable learners to express different communicative meaning” (Nunan, 2004, p.4). Thus, as noted in Chapter One, a focus on form (i.e., overtly drawing students’ attention to linguistic elements as they arise in lessons where the overriding focus on meaning, or communication), as distinct from forms, is equally important to TBLT (Long, 1991; Long & Robinson, 1998).

Despite the recognition about the importance of form, especially for pedagogic tasks, the way in which this should be included in task-based approaches is less clear. Ellis (2005, p. 243) does suggest a number of ways this might be done, for instance, through the manipulation of planning time as part of a pre-task. Specifically, in planning time learners are given the opportunity to plan or write what they are going to produce before carrying out the tasks. With respect to the current study, it should be noted that planning time is one element within the resource-depleting dimension, one of the two cognitive task complexity factors (Robinson, 2001, 2003, 2005). (See 2.4.1.1 for further discussion of planning time). As with tasks in general, the various dimensions of tasks are described differently by a number of researchers. The next section describes this in more detail.

2.3.3 Dimensions of Tasks

In order to understand tasks, different authors have categorized the dimensions or components in various ways. For example, Nunan (1989, 2004) proposes that six components should be taken into account in analyzing tasks: goals, input, procedures, settings, and teacher and learner roles. This is shown in Figure 1 below.

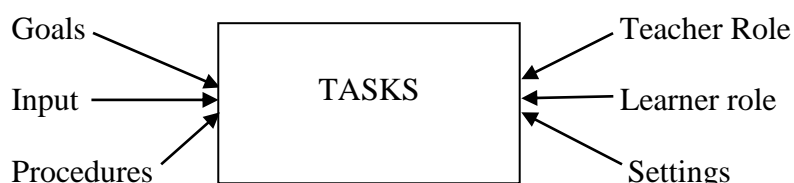


Figure 1: Components of Task (from Nunan, 1989, 2004).

According to Nunan, tasks are designed for learners to achieve goals. Specifically, learners need to be involved in activities with defined outcomes, which in turn may be communicative, affective and/or cognitive. Tasks also include input or materials in the form of either linguistic (e.g., newspapers, letters) or non-linguistic data (e.g., pictures stories, photographs). In addition, learners need to be provided with directions so that they can understand the procedures or activities they need to follow and/or engage in. Related to this, are the distinct roles that both the teacher and learners engage in when undertaking tasks. Finally, tasks include the setting in which the tasks are undertaken (e.g., a pair work or a group work situation).

In contrast, Ellis (2003, pp. 2-9) identifies six slightly different dimensions of a task, namely: i) its scope; ii) the perspective from which it is viewed; iii) its authenticity; iv) the linguistic skills required to perform it; v) the psychological processes involved in its performance; and, vi) its outcome.

The ‘scope’ of a task refers to whether it involves the use of language or has “language-free activities” (p. 2). This is in reference to Long’s (1985) definition of a task, that is, whether a task requires linguistic or non-linguistic data. The scope also concerns the focus of the task activities, such as whether they are primarily on meaning or on both meaning and form. Note, that there is some inconsistency in the

literature with respect to this as Skehan (1996) and Nunan (2004), among others, restrict the term *task* to activities which primarily focus on meaning.

According to Ellis (2003), task ‘perspective’ concerns whether they are designed from the learners’ or the designers’ point of view. Tasks, as he notes, have most commonly been designed from the designers’ point of view which is sometimes contradictory to the learners’ needs. Ellis gives as an example a task which is designed to encourage learners to focus on meaning-based engagement (a communicative language activity) and which may turn into a form-focused activity as learners sometimes try to “suit their own purposes” (p.5). This suggests that learners’ views about their ‘needs’ should play an important role in sequencing pedagogical tasks.

The third of Ellis’s (2003) dimensions, ‘authenticity’, refers to how closely the tasks reflect real-life activities. It should be noted, however, that ‘real-life activities’ do not always refer to those that occur in what Ellis (2003) calls “day-to-day living” (p. 6). In fact, authentic tasks can include activities that are only indirectly related to real-life such as describing pictures or comparing two similar things or pictures, because “the kind of language behavior they elicit correspond to the kind of the communicative behavior that arises from performing real-world tasks” (Ellis, 2003, p. 6). This, in turn, reflects the fourth dimension of tasks, linguistic skills. Ellis gives an example of this as being when learners are encouraged to compare two similar pictures and share information by asking questions and clarify meanings with each other. These sorts of activities are not only commonly found in real-life activities, but they do provide opportunities for learners to develop the target language.

In relation to this fourth dimension, Ellis does point out that many definitions of tasks have different points of view about what language skills may be involved in their performance. Ellis’s (2003) definition of tasks explicitly involves the engagement of productive and receptive skills, as well as those that are produced orally or as written texts.

The fifth dimension of a task, according to Ellis, relates to the psychological or cognitive process involved. Cognitive processes include such things as selecting,

reasoning, classifying, sequencing information, and transforming information. In this way this dimension closely aligns to Prabhu's definition of a task:

An activity which required learners to arrive at an outcome from given information through some process of thought, and which allowed teachers to control and regulate that process (1987, p. 24).

The role of cognitive factors in performing tasks has been outlined by Robinson (2001a; 2001b, 2003; 2005) in his discussions of task complexity. Specifically, and of particular relevance to the current study, he introduces resource-directing and resource-depleting dimensions. For example, a task which requires learners to explain many elements, as Robinson predicts, is expected to be more cognitively demanding than that with few elements to describe.

The last of Ellis's (2003) task dimensions is the 'outcome'. A task is not simply concerned with the use of language, but also requires the learners' successful performance in completing the tasks. Ellis (2003) gives the following example of a task outcome: "Learners can successfully describe all the differences and similarities between two pictures when they are asked to do so" (p. 9). Furthermore, according to Ellis (2005), the 'outcome' differs from the 'aim' of a task. The former refers to what the learners arrive at when they have completed the task, for example, a list of differences. The latter is concerned with the pedagogic purpose of the task, which is to elicit meaning-focused language use - either receptive or productive. Even so, the outcome is ultimately directed at achieving the aim of the task.

Perhaps as a consequence of the term *task* having a number of different definitions, it is often confused both in the literature and especially by practitioners with *exercises* or other activities. To address this issue, Skehan (1998, p. 95) has attempted to distinguish the difference between tasks and other language activities using the following five task criteria: 1) meaning is uppermost; 2) there is some communication problem to solve; 3) there is some sort of relationship that is comparable to real world activities; 4) task completion has some priority; and, 5) the assessment of the task is in terms of its outcome. He gives examples of classroom task-based activities such as, completing family trees, solving a riddle, leaving a message on someone's answering machine, and so on. In a different way, Bygate,

Cook, Iannou-Giorgiou, and Julian (2003) distinguish tasks from *exercises* in terms of integrative versus discrete activities. That is, *tasks* refer to activities which provide practice in using whole integrative skills in some way, while *exercises* refer to activities which provide practice in using parts of a skill, a new-sub skill, or a new piece of knowledge.

Similarly, Ellis (2009b, p. 223) regards a language learning activity as a *task* rather than an *exercise* when it meets the following criteria: 1) the primary focus is on meaning (i.e., learners should be mainly concerned with processing the semantic and pragmatic meaning of utterances); 2) there is some kind of a ‘gap’ (e.g., there is a need to convey information, to express an opinion or to infer meaning); 3) learners largely have to rely on their own resources (linguistic and non-linguistic) in order to complete the activity; 4) there is a clearly defined outcome other than use of language (i.e., the language serves as the means for achieving the outcome, not as an end in its own right). According to Ellis, these criteria distinguish between *tasks* and “a situational grammatical exercise” (p. 223).

In summary, therefore, tasks constitute language learning activities that are designed in such a way that learners actively engage cognitive processes in using the language as a means of communication, that is, they are meaning-focused activities. This will ultimately enhance the learner’s capacity to use the target language as a means of communication as commonly occurs in real-world activities.

2.4 Task-Based Research

In the last few years, research on TBLT within SLA has burgeoned as a consequence of an extensive use of this approach in language teaching worldwide. Further there is added utility of tasks within SLA research more generally because task types can be used to reflect learners’ language performance (Samuda & Bygate, 2008). Therefore, a frequent goal of task-based studies is “to establish whether the predictions made by the designers are borne out” (cf. Ellis, 2003, p.5).

So far, there have been two main approaches to task-based studies within SLA: the interactionist and the information-processing approaches. The former emerged from

the interaction hypothesis (Long, 1983, 1985, 1996), and the latter from the Cognitive Approach (Skehan, 1998) or the Cognition Hypothesis (Robinson, 2001a, 2001b, 2003, 2005, 2007a, 2007b, 2011). The two approaches differ from each other regarding the focus of the investigation, although both interrogate learner language production.

The interactionist approach focuses on those activities that trigger the promotion of interaction, such as negotiation of meaning and feedback (e.g., comprehension checks, clarification requests, and confirmation checks, and also implicit and explicit forms of feedback, such as recasts and metalinguistic comment). This approach is based on the argument that such interaction facilitates SLA (Long, 1983, 1985, 1996; Pica, Young, & Doughty, 1987; Gass, 1997; Gass, Mackey, & Pica, 1998). Long (1983, 1996) in his original and then updated Interaction Hypothesis, argues that acquisition takes place when learners obtain comprehensible input as a result of the opportunity to interact, especially when communication breakdown occurs. Task-based studies within the interactionist perspective are, therefore, characterized by interaction activities, with learners working in pairs or groups, and engaging in a number of different types of tasks such as making jigsaws, filling information gaps, or doing role plays, etc.

The initial studies into interaction involved tasks undertaken between L1 speakers of English and second language learners. This has been extended and extensive research has explored interactions between language learners in ESL contexts (e.g., Ellis, Basturkmen, and Loewen, 2001; Mackey, Kanganas, & Oliver, 2007), in EFL situations (e.g., Yufrizal, 2001; Iwashita, 2001; McDonough, 2004; Fernandez-Garcia, 2007; Fujii & Mackey, 2009), both in laboratories and classrooms (e.g., Gass, Mackey, & Ross-Feldman, 2005, 2011), and with adults (e.g., Storch, 2002; Nabei & Swain, 2010), children (e.g., Oliver, 1995, 1998, 2002; Mackey & Oliver, 2002; Mackey & Silver, 2005), and adults and children (e.g., Oliver, 2000; Mackey, Oliver, & Leeman, 2003).

However, studies undertaken within an interactionist perspective have been criticized for a number of reasons, including that much of the research was initially based on laboratory, rather than classroom settings (Foster, 1998). Given the body of research

that now exists, however, this criticism now has less validity (Gass, Mackey, & Ross-Feldman, 2005, 2011). Others, such as Skehan (2003), whilst supporting the key role of interaction, suggest that the focus of research should extend beyond interaction. In particular, he suggests that the focus of task research should be on understanding the psychological processes learners use when working on tasks (Skehan, 2003). As indicated above, this approach is called the Cognitive Approach or the Limited Capacity Model (Skehan, 1998) or the Cognition Hypothesis (Robinson, 2001a, 2001b, 2003; 2005; 2007a, 2007b). (Note that Skehan's (1998) Cognitive Approach (Limited Capacity Model) and Robinson's (2001a) Cognition Hypothesis are discussed in Section 2.4.1 below).

Ellis (2000) provides support for this position suggesting that the information-processing perspective, on which these models are based, "could be predictive and deterministic which mean that properties in a task will predispose or even induce learners to engage in certain types of language use and mental processing that are beneficial to acquisition" (p.197). This current study was informed by this perspective and is described in detail in the next section.

2.4.1 Tasks and information-processing

Task-based studies using the information-processing perspective predominantly adapt the theories of working memory system, in particular Levelt's (1989) model of speech production. This model was initially developed to account for L1 speech production. However, it has also been used to explain the effect of task complexity on L2 speech production in terms of complexity, accuracy, and fluency (CAF) (De Bot, 1992; Yuan & Ellis, 2003; Gilabert, 2005). For example, Gilabert (2005) attempted to use this model to explain the effect of task complexity manipulated along with planning time (+/- planning time) and present and past references (+/- here and now).

According to Levelt (1989), there are three main stages of speech production within the working memory system – conceptualization, formulation and, articulation. Firstly, a learner will conceptualize what she or he is planning to say before attempting to produce speech. This process results in the form of preverbal

messages. The conceptualizing process involves two sub-stages, macroplanning and microplanning. Macroplanning has to do with the “elaboration of some communicative goal into a series of sub-goals, and the retrieval of the information to be expressed in order to realize each of these sub-goals” (Levelt, 1989, p.11).

Microplanning assigns propositional shape to information chunks according to the speaker’s information perspective (the particular topic and focus). The process of both macroplanning and microplanning generates pre-verbal messages which take the form of non-linguistic input. The input is processed in the second stage, that is, formulation.

According to Levelt, the preverbal messages derived from the process of conceptualization are encoded in the formulation stage through the retrieval of lexical items stored in the mental lexicon. These lexical items consist of two components: Lemma and Lexeme. Lemma contains semantic and syntactic information, while lexeme deals with the information about its morphological and phonological properties. The process of retrieving lexical items from the lexicon functions to trigger syntactic building procedures for grammatical encoding to take place. This process results in what Levelt (1989) refers to as “surface structure”, which consists of “ordered strings of lemmas grouped in phrases and sub-phrases of various kinds” (p.11). These phrases and sub-phrases are then processed phonologically, which results in a phonetic or articulatory plan, that is, “an internal representation of how the planned utterance should be articulated”, which Levelt refers to as “internal speech” (p. 12). Finally, the internal speech developed in the formulation stage is executed in the form of overt speech production. In turn, learner speech performance can then be measured in terms of how well it is produced – most often in terms of its complexity, accuracy and fluency (i.e., CAF).

Therefore, based on Levelt’s theory of information-processing, task-based studies have been undertaken investigating the differential effects of task factors on language performance in terms of CAF. For example, studies have investigated the effect on CAF on manipulating planning time (Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Yuan and Ellis, 2003, 2005; Philp, Oliver, & Mackey, 2006; Foster & Tavakoli, 2009; Ahmadian & Tavakoli, 2010), number of elements (Robinson, 2001a; Michel et al., 2007; Kim, 2009), here and now (present and past

references (Ishikawa, 2007), number of elements and reasoning demand (Kuiken & Vedder, 2007, 2008; Gilabert, et al., 2009) and, planning time and present and past references (Gilabert, 2005).

To date, however, most task-based studies undertaken from an information-processing perspective have been conducted predominantly using monologic tasks (one learner's production) along with the manipulation of only one task factor (e.g., planning time) rather than with dialogic tasks (i.e., two learners interacting) and manipulating more than one task factor at a time (e.g., planning time and tense) (e.g., Robinson, 2001a, 2007a; Yuan and Ellis, 2003; Gilabert 2005; Foster & Tavakoli, 2009; Ahmadian & Tavakoli, 2010; Tavakoli & Foster, 2011; Kormos and Trebits, 2012). Although Gilabert (2005) did investigate two task-complexity factors simultaneously, namely present and past preference (+/- here and now) and planning time (+/- planning time), once more only monologic tasks were used (specifically learners were asked to narrate stories rather than discussing them with a friend). Further, although Michel, Kuiken, and Vedder (2007) compared monologic and dialogic tasks, they did so manipulating only one task factor, that is, the number of elements (+/- few elements). Bell (2003) also compared monologic and dialogic tasks, in particular measuring just one element, namely the fluency of performance of six young adult learners of English as a foreign language. The results show that the majority of learners produced more fluent oral production in terms of speech rate for the dialogic tasks than the monologic ones. In short, few studies informed by the information-processing perspective have been conducted along two dimensions simultaneously (e.g., planning time and the number of elements) using dialogic tasks. The current study did this.

As the current study is concerned with task-based approaches informed by the information-processing perspective, this chapter will detail Skehan's (1998) Limited Capacity Model/Cognitive Approach and Robinson's (2001a) Cognition Hypothesis. This includes a discussion of those CAF constructs traditionally used to measure learners' language performance.

2.4.2 Skehan's Limited Capacity Model/Cognitive Approach

Skehan (1996, 1998) describes how human beings have limited attentional capacity and that this impacts on language performance, particularly when learners perform difficult tasks. Skehan and Foster (2001) account for task difficulty from a cognitive perspective, suggesting that:

Task difficulty has to do with the amount of attention the task demands from participants. Difficult tasks require more attention than easy tasks (p. 196).

Thus, according to Skehan, “task difficulty” (or “task complexity”), involves heightened cognitive engagement. Note, that this differs from the type of difficulty Robinson (2001) refers to as “learner affective factors” (see discussion of this in 2.3.3). Further, Skehan predicts that tasks which are made more difficult (more cognitively engaging), will decrease learners’ L2 performance in terms of complexity, accuracy, and fluency (CAF) because their attentional resources are forced to primarily focus on meaning rather than on form.

On this basis, and drawing on Candlin’s (1987) and Nunan’s (1989) analyses of tasks sequencing, Skehan (1996, 1998) suggests that tasks should be developed on the basis of the three criteria, namely “the language required, the thinking required, and the performance conditions for a task.” (p. 99). He describes these three criteria as ‘code complexity’, ‘cognitive complexity’, and ‘communicative stress’ respectively, and maintains that they can be used for assessing task difficulty. The constituent parts of these criteria are, as follows:

1. **Code Complexity:** According to Skehan (1996, 1998), this refers to the areas of syntactic and lexical difficulty (linguistic demands imposed on tasks) and include: linguistic complexity and variety; vocabulary load and variety; redundancy; and, density.
2. **Cognitive Complexity** - “the content of what is said”, that is, the process of the conceptualization as explained in Levelt’s (1989) information-processing model. This stage has two distinct aspects, cognitive familiarity and cognitive processing (p. 99).

- a. *Cognitive Familiarity* is concerned with “the amount of on-line computation that is required while doing a task, and highlights the extent to which the learner has to actively think through task content” (p. 99), that is, the familiarity with the topic and its predictability, the familiarity with the discourse genre, and the familiarity with the task;
 - b. Cognitive Processing involves “the extent to which the task draws on ready-made or pre-packaged solutions” (p.99), that is, the materials that are accessed relevant to schematic knowledge and includes information organization, amount of ‘computation’, clarity and sufficiency of information given, and, information type.
3. **Communicative Stress** relates to how the tasks are performed (e.g., under the planned or the unplanned conditions). Skehan argues that communicative stress is not directly related to code and meaning, but it can lead to the “pressure of communication” (p. 99), that make tasks difficult to perform, and includes such things as time limits and time pressure, speed of presentation, number of participants, length of texts used, type of response, opportunities to control interaction.

Skehan (1996, 1998) argues that together these three aspects of task difficulty (i.e., code complexity, cognitive complexity, communicative stress) impact on learners’ L2 performance in terms of CAF. Further, he argues that if these tasks are too difficult (i.e., requiring increased cognitive engagement) they will decrease all aspects of CAF. Conversely, he predicts that tasks which are too easy (cognitively less demanding), might not significantly facilitate learners to develop their L2 performance in terms of CAF. Therefore, Skehan suggests that the degree of task difficulty should be moderated so that learners have some chance of balanced attention.

2.4.3 Robinson’s Cognition Hypothesis

The Cognition Hypothesis (Robinson, 2001a, 2001b, 2003, 2005, 2007a, 2011) claims that pedagogic tasks should be designed and sequenced on the basis of task complexity, specifically in terms of the manipulation of cognitive factors. Robinson distinguishes between the terms task complexity (cognitive factors) and task

difficulty (learner factors), which were previously used interchangeably (e.g., Skehan, 1996, 1998; Skehan & Foster, 2001). He also distinguishes task complexity and task difficulty from task conditions (interactive factors). Therefore, Robinson (2001a, 2001b, 2003, 2005, 2007a) proposes the Triadic Componential Framework composed of three aspects, task complexity (cognitive factors), task conditions (interactive factors), and task difficulty (learner factors). Robinson argues that these three factors influence learners' L2 performance in terms of CAF. The Triadic Componential Framework has been widely adopted as the construct for task-based studies. The components of Robinson's Triadic Framework can be seen in Figure 2 below.

Task complexity (Cognitive factors)	Task conditions (Interactive factors)	Task difficulty (Learner factors)
a) resource-directing e.g., +/- few elements +/- here-and now +/- no reasoning demands	a) participation variable e.g., one-way/two-way convergent/divergent open/closed	a) affective variables e.g., motivation anxiety confidence
b) resource-depleting e.g., +/- planning +/- single task +/- prior knowledge	b) participant variables e.g., gender familiarity power/solidarity	b) ability variables e.g., aptitude proficiency intelligence
Sequencing criteria ----- Prospective decisions about task unit		Methodological criteria on-line decision about pairs and groups

Figure 2: Robinson's Triadic Componential Framework

Each component of the Triadic Framework is described in more detail in the following sections. The factors of task complexity (cognitive factors) both the resource-directing and the resource-dispersing dimensions, are described in detail below.

i) Task Complexity (Cognitive Factors)

As previously stated, Robinson (2001a, 2001b, 2005, 2007a, 2011) argues that the three factors (i.e., cognitive, interactive and learner factors) affect learners' L2 performance. Further, Robinson suggests that pedagogical tasks should be sequenced

only on the basis of cognitive factors (or what he labels *task complexity*) because learner factors cannot be used to predict task difficulty in advance.

Robinson (2001a) maintains that task complexity, is

the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner. These differences in information processing demands, resulting from design characteristics, are relatively fixed and invariant (p. 28).

With particular relevance to the current study, Robinson outlines the two dimensions of *task complexity* as the resource-directing and resource-dispersing (Robinson, 2001a, 2001b, 2003, 2005, 2007a, 2007b). These dimensions constitute cognitive factors that can be manipulated to increase or lessen learners' cognitive engagement when learners are performing tasks. He hypothesizes that the manipulation of *task complexity*, either the resource-directing or the resource-depleting dimensions, will facilitate second language learning. In addition, Robinson argues that the resource-directing, the so-called developmental dimension, refers to cognitive/conceptual demands requiring attention and working memory that directs learners to focus on linguistic form. The resource-directing dimension includes +/– here and now, +/– few elements, +/– no reasoning demands. In this case tasks represented by “+” refer to simple, present, or less components, while complex tasks symbolized by “–” mean complex, absent, or more components. Therefore, tasks along the resource-directing dimension are represented by + here and now, + few elements, and + no reasoning demands. On the other hand, the complex tasks are coded as –here and now, – few elements, and – no reasoning demands.

Robinson (2001a, 2001b, 2003, 2005, 2007a, 2007b, 2011) argues that tasks in the resource-directing dimension are expected to be easier for learners to perform as they are less cognitively demanding than the complex tasks. On the one hand, Robinson predicts that tasks with increasing complexity along the resource-directing dimensions should be more difficult as they involve more cognitive engagement. For example, according to Robinson, tasks that require a simple description of events happening in the present and in a shared context (+ *here and now*), with few elements to be described and distinguished (+ *few elements*), and not requiring

reasons to be given (+ *no reasons*) are less cognitively demanding than tasks that happened elsewhere and in the past (– *here and now/there and then*), that include many elements that need to be described and distinguished, and that also require a lot of reasons to be given (– *no reasoning*).

The other dimension of task complexity is that labeled as the resource-depleting/dispersing. This refers to “performative or procedural demands” that require learner’s attentional and memory resources, but at the same time do not direct learners to any particular linguistic form. This dimension includes +/–planning, +/–single task, and +/–prior knowledge. (Again, simple tasks are represented by ‘+’, while complex ones are symbolized by ‘–’). Simple tasks along this dimension are coded as + planning time, + prior knowledge, + single task, whereas complex ones are represented by – planning time, – prior knowledge, – single task. Robinson (2001a, 2003, 2005, 2007) argues that tasks which are made complex along this dimension will also require more engagement in terms of attention and working memory. Therefore, the tasks where planning time and prior knowledge are available and require a single activity, + planning, + prior knowledge, + single task, will be less cognitively demanding than tasks that are without planning time and prior knowledge in two-way tasks, –planning, – prior knowledge, – single tasks.

Robinson (2001, 2005) argues that tasks which are made more complex (increasing complexity) along the resource-directing dimension will result in increased conceptual or functional requirements. As a consequence, this will decrease learner fluency, but will increase the accuracy and complexity of the language production. In contrast, Robinson claims that tasks which are made more complex (increasing complexity) along the resource-depleting dimension (e.g., –*planning time*) will decrease all aspects of CAF because learners will find it difficult to access their current repertoires of L2 knowledge. Consequently, increasing task complexity along the resource-directing dimensions can be expected to have a positive effect on learners’ language production when the task is simultaneously simpler along resource-dispersing/depleting dimensions, (i.e., – few elements, + planning time), compared to when it is complex along both sides of the dimensions (i.e., – few elements, – planning time).

In this way Robinson (2005) disagrees, at least in part, with Skehan's (1998) Limited Capacity Model which states that increasing task complexity along the resource-directing dimension will have a negative effect on complexity, accuracy, and fluency (CAF). Robinson, (2005) argues that increasing complexity along the resource-directing dimension will only decrease fluency. In all other respects, Robinson agrees with Skehan's hypothesis that increasing task complexity along the resource-depleting dimensions will decrease the CAF of learner language production. In particular, he argues that learners will find it difficult to access their existing repertoire of L2 knowledge if there is insufficient time and learner background knowledge is limited. This information is summarized and presented in Table 1, below.

Table 1: Configurations of the binary code (+/–) of task difficulty (Cognitive factors)

The Resource-directing dimensions		The Resource-dispersing dimensions	
Code	Explanation	Code	Explanation
+ <i>few elements</i>	has few elements	+ <i>planning time</i>	has planning time
– <i>few elements</i>	has many elements	– <i>planning time</i>	has no planning time
+/– <i>few elements</i>	has either few or many elements	+/– <i>planning time</i>	has either planning or no planning time
+ <i>here and now</i>	uses present tense	+ single task	single tasks
– <i>here and now</i>	uses past references (there and then)	– <i>single task</i>	dual tasks
+/– <i>here and now</i>	uses either present or past references	+/– <i>single task</i>	either single or dual tasks
+ <i>no reasoning demand</i>	requires no reasoning demand	+ <i>prior knowledge</i>	has background knowledge/schemata
– <i>no reasoning demand</i>	requires reasoning demand	– <i>prior knowledge</i>	has no background knowledge/schemata
+/– <i>no reasoning demand</i>	requires either no reasoning demand or reasoning demand	+/– <i>prior knowledge</i>	has either schemata or no schemata

Planning time is one factor within the resource-depleting dimension that has long been acknowledged as an important part in the process of oral production. As Faerch and Kasper (1983) note, there are two phases of oral production, planning

and execution. In the planning phase, the speaker first attempts to scan their linguistic repertoire and then retrieves linguistic rules and lexical items for the purpose of reaching a communicative goal, and finally the two activities (the examination of linguistic repertoire and the retrieval of linguistic rules) are integrated (Faerch and Kasper, 1986). Planning is argued to be an effective way to reduce the cognitive load of demanding activities (Crookes, 1989; Foster & Skehan, 1996; Skehan, 1996; Ellis, 2003). Further, it does seem that planning has an important role in facilitating learners to improve the CAF of their language performance (Yuan & Ellis, 2003, 2005; Ellis, 2005; Gilabert, 2005; Philp, Oliver, & Mackey, 2006; Gilabert, 2007; Ahmadian & Tavakoli, 2010).

Ellis (2005, p. 3) claims that “planning is essentially a problem solving activity; it involves deciding what linguistic devices need to be selected in order to affect the audience in the desired way.” Ellis (2005) further argues that:

Planning is seen as a means of helping learners overcome the limitations in the capacity of their working memory. Providing learners with the opportunity to plan a task performance constitutes a means of achieving a focus-on-form pedagogically. It mitigates the limitations of their working memory by allowing learners the cognitive windows needed to attend to form while they are primarily to map form onto meaning by accessing linguistic knowledge that is not yet automatized (pp. 9-10).

Ellis (2005, p. 4) divides planning into two principle types. *Pre-task planning* and *within-task planning* (see Figure 3 below). The former refers to the planning that takes place before and the latter during the performance of the task. Pre-task planning is further divided into rehearsal and strategic planning. Rehearsal refers to the opportunities to perform or repeat the tasks before the ‘main performance’ whereas strategic planning entails preparation to perform the tasks and express the content. Specifically, in pre-task planning, learners have access to the actual task materials and this distinguishes strategic planning from other pre-activity (e.g., brainstorming content). On the other hand, within-task planning refers to pressured or unpressured planning of the task performance, such as can be achieved by manipulating the planning time. Quite early, Ochs (1979) described how “planned language use” can result in more complex language production and target like forms.

On the other hand, in pressured performance learners are engaged in rapid planning, resulting in ‘unplanned language use’ which, in turn, may lead to the production of non-standard language.

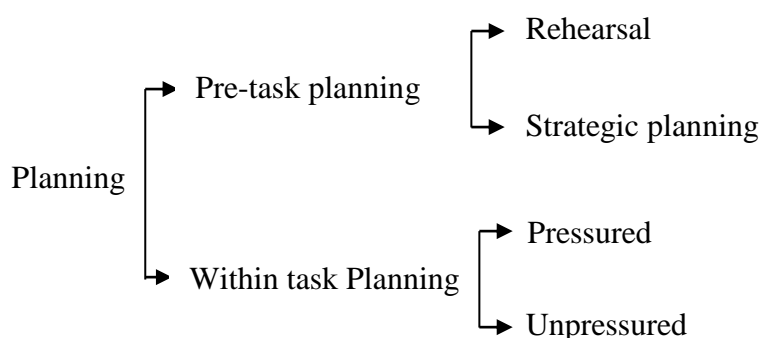


Figure 3: Ellis’ (2005) model of task-based planning

A number of studies of tasks investigating the resource-depleting dimension (i.e., manipulating *+/- planning time*) and using measures of CAF for analysis have been conducted (e.g., Foster & Skehan, 1996; Mehnert, 1998; Ortega, 1999; Philp, Oliver, & Mackey, 2006; Yuan & Ellis, 2003, 2005; Gilabert, 2007; Ahmadian & Tavakoli, 2010). Generally, the results show positive effects on the learners’ productions. For example, in one of the earlier studies, Foster and Skehan (1996) explored the influence of planning time and task types on second language performance in terms of CAF. The study was conducted with 32 pre-intermediate-level adult EFL learners of English. In their study three different tasks were employed, personal information exchange, narrative and decision-making, and each type of task was manipulated under the three planning conditions, unplanned, undetailed, but planned (i.e., ten minutes planning without guidance), and detailed planning (i.e., ten minutes planning with guidance about how to plan what to say). The results showed that ten minute detailed planning time resulted in more fluent speech and more accurate sentences (as measured by clauses per C-units). Thus the results suggest that planning has an effect on fluency and complexity. However, more complex relationships were found between planning and accuracy, with most accurate performance produced by the less detailed planners. In addition, interaction effects were found between task types and planning conditions. Specifically, the effects of planning were greater with the narrative and decision-making tasks than with the personal information exchange tasks.

Similarly, Mehnert (1998) investigated the effect of planning time (i.e., no planning time, one minute, five minutes, and ten minutes) on four different groups of learners. The results showed that the task with ten minute planning time generated significantly more fluent, more accurate, and more lexically dense oral production, but no statistically significant difference was found for syntactic complexity. Similarly, another planning time study conducted by Ortega (1999) found that learners were able to produce significantly more fluent and complex language when planning time was provided. One explanation for this is that giving learners the opportunity to plan before they perform tasks enhances their attention to form. Ortega also makes the point that learners appear to employ a wide variety of cognitive and metacognitive strategies in their oral production and seem to monitor their output when given opportunity to plan. On this basis Ortega argues that planning may lessen the cognitive load and free up attentional resources needed to accomplish the task.

Yuan and Ellis (2003) compared task complexity manipulated under planning, pre-task planning, and on-line planning with 42 Chinese students learning English as a second language. The results suggested that pre-task planning improves grammatical complexity while on-line planning positively influenced accuracy and grammatical complexity. Pre-task planning also produced more fluent and lexically varied language than on-line planning. Therefore, the results indicate that there is a close relationship between planning and L2 oral production.

In a different study, Yuan and Ellis (2005) compared tasks using within-task planning, for both oral and written tasks, with 42 Chinese learners of English. The tasks employed two conditions, pressured planning (PP) and careful planning (CP). The learners were divided into three groups and randomly assigned to one of three groups. The students in each group were asked to perform, in turn, the two types of tasks using both speaking and writing production. The result showed CP resulted in greater syntactical complexity and accuracy than PP, but there was no statistically significant effect on fluency or lexical variety. With regard to modality, speaking proved more fluent than writing, but writing was characterized by greater syntactical and lexical complexity and also increased accuracy.

Gilabert (2007a) examined the effect of planning time (+/- *planning time*) and tense (+/-*here-and-now*) on learner oral production when undertaking tasks (as measured by CAF). The study was undertaken with 48 young Spanish students with low-intermediate proficiency levels of English. They were asked to narrate four wordless comic strips in four different conditions (i.e., + *planning time* and + *here-and-now*; + *planning time* and -*here-and-now*; - *planning time* and + *here-and-now*; and, - *planning time* and -*here-and-now*). The results indicated that these conditions had an effect on fluency. Both simple *here-and-now* and complex *there-and-then* tasks generated a significantly higher speech rate when performed ten minutes of planning time. Both the planned *here-and-now* and *there-and-then* tasks triggered significantly more fluent oral production than the unplanned *here-and now* and *there-and-then* tasks. There were also significant differences in complexity in learner vocabulary between the planned *here-and-now* and *there-and-then* tasks than the unplanned conditions. However, no significant differences were found on either structural complexity or accuracy between both the planned *here-and-now* and the planned *there-and-then* tasks and both the unplanned ones.

In short, the empirical studies have shown that the tasks performed within the planning condition, particularly ten minute planning time, generated improved language performance in terms of CAF than the unplanned condition, particularly with the monologic task conditions. These results informed the methodology used in the current study.

As previously described above the resource-directing dimension of task includes three components: +/-*few elements*, +/- *reasoning demand*, and +/-*here and now*. Among these three components, the manipulation of a number of elements is regarded to be more inclusive than the other two components (+/- *reasoning demand*), and (+/-*here and now*). This is because tasks which are manipulated according to number of elements (+/-*few elements*) are expected to involve the other two components of the resource-directing dimension, namely, giving reasons (+/-*reasoning demands*) and using present or past references (+/-*here and now*).

Previous studies investigating the number of elements have been conducted by such researchers as Robinson (2001a), Kuiken and Vedder (2007), Michel et al. (2007),

Gilabert (2007b). Robinson (2001a), for example, conducted a study using monologic tasks where he simultaneously combined prior knowledge and the number of elements. The aim of the study was to test his Cognition Hypothesis (2001a), namely, that a task made more complex will decrease learner fluency, but generate more accurate and complex language production. In the study one task the learners performed was simple and included a map of an area with which the learners were already familiar (+ *prior knowledge*) and contained only a few elements (+ *few elements*). On the other hand, the complex task, also a map task, but included many elements (– *few elements*) of an area with which the students were not familiar (– *prior knowledge*). Students were asked to give directions to another student who had to draw a route on an empty map. The results showed that the complex task resulted in significantly less fluent oral production, but higher lexical complexity than the simple task. However, the complex task did not affect either accuracy or syntactic complexity. According to Robinson (2001a), the nature of interactive tasks, with many turn-takings and interruptions, may “mitigate learners’ attempts at using structurally complex language” (p. 36).

Kuiken and Vedder’s (2007) study examined two task factors, specifically the number of elements (+/– few elements) and the reasoning demands (+/– no reasoning demands) within the resource-directing dimension. The subjects of the study were seventy-six adult learners of French with different proficiency levels. The findings suggested that increasing task complexity along the resource directing dimension resulted in more accurate language production. It should also be noted that no effect for task complexity and proficiency level was observed.

Further support for Robinson’s Cognition Hypothesis (i.e., that cognitively more demanding tasks result in more accurate written output than cognitively less demanding tasks) was found in the next study conducted by Kuiken and Vedder (2008). In this research they examined the effect of complexity on the written output of 91 Dutch learners of Italian and 76 Dutch learners of French of different proficiency levels (low and high). Once more, no effect for text complexity and proficiency level was found. Similarly, Gilabert (2007b), who examined task along one dimension (resource-directing) by manipulating three types of tasks (+/– here and now, +/– few elements, and +/– few reasoning), found no effect for proficiency

level. However, rather than measuring output according to CAF, in this study learner self-repair behavior was explored. The results showed that the complexity of three types of tasks did have a positive effect on self-repair.

Unlike the previous studies, Michel et al. (2007) examined the influence of task complexity, according to the number of factors (+/- few elements) using both monologic and dialogic tasks (most other studies used only monologic or dialogic tasks, but mostly the former). This study was conducted with learners studying Dutch as an L2. The results showed that dialogic tasks led to more accurate oral production in terms of the number of errors, omissions, and the ratio of self-repairs to errors, but generated lower accuracy in terms of the percentage of self-repairs. In contrast, the dialogic task generated a higher percentage of lexical richness (using Guiraud's Index – note this is described in Chapter Three), but produced lower syntactic complexity. The dialogic tasks resulted in more fluent oral production than the monologic tasks. Furthermore, the results showed that the complex dialogic task generated higher accuracy in terms of the number of errors, omissions, and the ratio of self-repairs to errors than the simple dialogic task, but resulted in a lower percentage of self-repairs. Syntactic complexity was lower in complex dialogic tasks, but lexical complexity increased in the complex dialogic task. In contrast, the simple dialogic task generated more fluent oral production (as measured by Unpruned Fluency Speech Rate A and Pruned Speech Rate B) than the complex dialogic task. In this respect, the findings only partially support the Cognition Hypothesis.

ii) Task conditions - Interactive Factors

Robinson (2001, 2003, 2005) specifies two types of task conditions or interactive factors, namely participation factors (interactional demands) and participant factors (interactant demands).

Participation factors include whether tasks are one-way (i.e., one learner holds all the information) or two-way (i.e., both learners hold parts of the information), convergent (i.e., one solution to the task) or divergent/open (i.e., many possible solutions). Convergent and divergent tasks refer to the goal-orientation of the task (Duff, 1986). For example, Duff gives an example of a convergent task, the "Desert

Island” task. In this task, learners must agree on a limited number of objects to take to the island and must, therefore, work together to find an acceptable solution. In contrast, a divergent or open task would be a debate about a controversial issue to which each learner in a group or in the class can contribute ideas. Another aspect of the task condition is flow and distribution of information among the learners, in particular whether they are monologic or dialogic tasks. As this has particular relevance in the current study, this is described in detail.

As indicated, information can be exchanged between the learners either in a one-way or a two-way flow. In monologic (one-way) tasks, information flows in one direction. On the other hand, in dialogic (two-way, interactive) tasks, information flows in more than one direction and each member of a dyad or a group has part of the information, which she or he must share with the others for the task to be completed successfully (Pica et al., 1993; Gilabert, 2005, p. 153-154).

When students perform monologic tasks, their role is to provide information to their partners, whereas the role of the other students in the dyads is simply to be recipients of the information, although they may ask questions to establish the meaning. Although monologic tasks can be interactive, the roles are not balanced as the information predominantly flows in one direction. On the other hand, in dialogic tasks learners are expected to be actively involved in sharing ideas as the information flows in two-ways. It is claimed that “through dialogic interaction, teachers can provide learners with effective assistance that will enable them to perform at higher levels than they would otherwise” (Anton, 1999, p.304). Ellis (2003, p. 177) also suggests that “dialogic discourse is better equipped to identify what a learner can and cannot do without assistance. It serves to create the intersubjectivity that enables verbal interaction to mediate learning”.

Robinson (2003, p.64) predicts that complex dialogic tasks should result in less fluent, but more accurate language production, but that simple dialogic tasks should generate more fluent language production, but they will decrease in accuracy. Robinson, (2003, 2005, 2007) further speculates that interactive complex tasks are expected to trigger learners to produce comprehension checks and clarification requests, which can decrease syntactic complexity.

As indicated above, when Michel et al. (2007) compared task complexity using monologic (one-way) and dialogic (two-way) tasks, manipulating the number of elements (+/- *few elements*) in L2 Dutch, the results showed that increasing task complexity resulted in more accurate, but less fluent oral production. Furthermore, the dialogic tasks triggered more accurate and more fluent oral production, but the production was structurally less complex. However, the interaction of task complexity and task conditions showed that the tasks performed under the monologic condition generated more accurate oral production. The results of the study partially supported the cognition hypothesis. Clearly there is a need for further research comparing monologic and dialogic tasks, informed by previous research within the resource-depleting dimension. It was one purpose of the current research to do this.

Participant factors also include such things as gender, familiarity, power or solidarity, and so on. As with various terms in this research field, there is some variability in the definitions. For example, the factor labeled familiarity has been interpreted differently by different researchers (Skehan, 1998; Bygate, 1999; Robinson, 2001a). Skehan (1998) interprets familiarity in terms of cognitive familiarity which includes: familiarity of topic, familiarity of discourse genre, and familiarity of task. Bygate (1999) refers to familiarity in terms of task repetition. Bygate argues that similar tasks which are performed repeatedly lead to better language performance in terms of CAF because repetition provides learners with “the time and awareness to shift attention from message content to the selection and monitoring of appropriate language” (p.41). On the other hand, Robinson’s (2001a) interpretation of task familiarity refers to the familiarity with the content of the tasks. Robinson gives an example of familiarity with a route marked on a map. According to Robinson, learners are more likely to find it easier to perform the tasks if they are familiar with the content or the topic of the tasks. On the other hand, they might find difficulties when the content or the topic of the task is unfamiliar to them.

iii) Task Difficulty – Learner Factors

Robinson (2001a, 2001b, 2003) distinguishes task complexity and task conditions from task difficulty (learner factors). Whereas the former refer to the manipulation of factors that can be imposed on tasks and manipulated in advance, the latter refers to task difficulty based on learner perceptions which emanate from what they bring to the task. In this respect task difficulty includes two aspects: 1) affective variables (i.e., motivation, anxiety, and confidence), and, 2) ability variables (i.e., aptitude, proficiency, and intelligence). As such, task difficulty resulting from learner factors cannot be easily predicted in advance. Further, Robinson (2001b, p. 31) argues that complexity and difficulty do not always have a fixed relationship to each other for two reasons. First, learners with different aptitudes may have different perceptions of the task difficulty. For example, the same task can be regarded as difficult by one learner, but easy by another learner. This, as Robinson argues, is as a result of inherent ability differentials between them, such as differences in the limits of their attentional, memory, and reasoning resource pools. However, the differences in learners' inherent abilities can also be affected by other factors such as motivation. According to Robinson, a learner with higher motivation, but low ability, for example, may struggle to expand "the resource pool currently available to meet the demands of the task", compared to a highly intelligent learner with lower motivation. Thus, it is necessary to explore learners' perceptions of task difficulty as a way to help explain different language performances on tasks. It is one aim of the current research to do this.

To ascertain learner perception Robinson (2001a, p. 41) specifies five categories: Level of difficulty, stress, confidence, interest, and motivation. In turn, he has developed five questions representing each of these categories which can be rated, by learners, on a nine-point Likert scale. In this way the questions are designed to investigate learners' perceptions of task difficulty based on both affective and ability factors and they are asked shortly after they have finished performing the tasks. The questions are as follows:

1. I thought the task was easy 1 2 3 4 5 6 7 8 9 I thought the task was hard.
2. I felt relaxed doing this task 1 2 3 4 5 6 7 8 9 I felt frustrated doing this task.
3. I didn't do well on this task 1 2 3 4 5 6 7 8 9 I did well on this task.

4. The task was not interesting 1 2 3 4 5 6 7 8 9 The task was interesting.
5. I don't want to do the task like this 1 2 3 4 5 6 7 8 9 I want to do the task like this.

(Robinson, 2001a, p.41).

As learner perception is one focus of the current study, these questions were used to inform the development of the data collection materials. Given the importance of 'perception of task difficulty' to the current study, this is discussed in detail in the next section.

2.5 Perceptions of Task Difficulty

As previously described, task-based studies undertaken from an information-perspective have addressed not only the impact of task conditions on CAF, but also the learner's perception of task difficulty. Although the learner's performance on CAF might be regarded as an indication of the extent to which a learner will find it easy or difficult to perform the task, CAF does not reflect a learner's affective and ability variables. Therefore, Tavakoli (2009, p. 1) suggests that learners' perceptions of task difficulty is necessary "to broaden the current understanding of task difficulty"

Studies to investigate learners' perceptions of task difficulty have been conducted by many researchers over a number of years (e.g., Candlin, 1987; Nunan & Keobke, 1995; Robinson, 2001a, 2007b; Tavakoli, 2009). Nunan and Keobke (1995), for example, investigated learners' perception of task difficulty with 35 Chinese learners of English using reading, listening, and speaking tasks. The findings showed that the learners found it difficult to perform the tasks due to three main factors - lack of familiarity with the topic, confusion over task purpose, and, difficulties regarding cultural knowledge.

In one of Robinson's studies (2001a) he explored learners' affective perceptions of the task difficulty, in addition to the learners' language performance in terms of CAF. The learners were asked to rate their responses to five perception questions (as outlined in the previous section). The results showed that the manipulation of task complexity corresponded to the learners' perception of task difficulty. That is, the

learners regarded the complex tasks as being more difficult and stressful than the simple tasks, and they also lacked confidence to perform the complex tasks. However, there was no difference in the learners' interest and motivation according to task complexity. Interestingly, fluency correlated with learners' perceptions of their ability to complete the task in both the simple and the complex versions of the tasks.

Drawing on Skehan's (1998) scheme for task difficulty and Robinson's (2001b) triadic componential framework (i.e., task complexity, task conditions, and task difficulty), Tavakoli (2009) investigated perceptions of task difficulty from both learners' and teachers' perspectives. Specifically, he investigated the degree of fit between teachers' and students' perceptions of task difficulty. The results showed that the teachers and the learners were in agreement on the six aspects underlying tasks difficulty in terms of cognitive demand, linguistic demand, and clarity of pictures/story, amount of information, task structure, and affective factors. The cognitive demand was the most frequently mentioned aspect amongst the six aspects underlying task difficulty. In short, the results largely confirmed the principles of the cognitive models of task difficulty proposed by both Skehan and Robinson.

Although studies to investigate of learner perceptions of task difficulty have been numerous, few studies have been conducted that explore learners' perceptions of task difficulty in depth. It was one aim of the current study to do this.

2.6 Measuring Complexity, Accuracy, and Fluency (CAF)

As indicated, in TBLT research CAF are regarded as the manifestation of learners' language performance. According to Housen and Kuiken (2009, p. 462), CAF emerge "as principal epiphenomena of the psycholinguistic mechanisms and process underlying the acquisition, representation and processing L2 knowledge." As these measures are integral to the current study, they are described in detail below.

In early work in this area, Wolfe-Quintero, Inagaki, and Kim (1998) described complexity and accuracy as mainly being concerned with the current level of the learner's interlanguage knowledge (i.e., L2 knowledge of syntactic and lexical rules).

Specifically, they defined complexity as “the scope of expanding or restructured second language knowledge” and accuracy is viewed as ‘the conformity of second language knowledge to target language norms” (p. 4). In other words, complexity and accuracy are regarded as the representations of learners’ L2 knowledge of the target language. In contrast, fluency constitutes a “function of the control in assessing the L2 knowledge, with control improving as the learner automates the process of gaining access” (p.4). As such fluency is reflected in the speech rate and the ease of access to the repertoires of L2 knowledge to express the ideas.

Since this early work various aspects of CAF have been used to measure learners’ L2 language performance. Despite this variability, there are common features or aspects of CAF as discussed in the following sections.

2.6.1 Complexity

Complexity is regarded as the most problematic measure amongst the three CAF measures (Pallotti, 2009; Housen & Fuiken, 2009). According to Pallotti, (2009, p. 592) complexity is considered problematic because it interchangeably refers to the properties of tasks and language performance. Similarly, Housen and Fuiken (2009 p. 463) regard complexity as “the most complex, ambiguous, and least understood dimension of the CAF measures” for two reasons. Firstly, the term ‘complexity’ can refer to two different properties, the properties of language tasks (the so-called task complexity), and the properties of L2 performance and proficiency, also known as L2 complexity. Secondly, the term “L2 complexity” has also been interpreted in two different ways, that is, as cognitive complexity and linguistic complexity. Although both types of complexity essentially refer to properties of language features, they have been defined from different perspectives. The former is defined from the perspective of the L2 learner-user, while the latter is defined from the perspective of the L2 system or the L2 features. In addition, cognitive complexity is regarded as a broader notion than linguistic complexity because “cognitive complexity of an L2 feature is a variable property which is determined both by subjective, learner factors such as aptitude, motivation, etc., as well as objective factors such as its inherent linguistic complexity” (Housen & Fuiken, 2009, p. 463).

Skehan and 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 taken to correlate with a greater likelihood of restructuring, that is, change and development in the interlanguage system (p. 96-97).

According to Ellis and Barkhuizen (2005, p. 139), complexity is “the extent to which learners produce elaborated language”. From the definitions of complexity by Skehan and Foster (1999) and Ellis and Barkhuizen (2005), it appears that complexity is concerned with learners’ current state of L2 interlanguage knowledge. That is, complexity refers to “the scope of expanding or restructured second language knowledge” as described by Wolfe-Quintero et al. (1998, p. 4).

Given the range of definitions, it is not surprising that complexity has been measured differently by different researchers. Despite complexity having been measured in different ways, there are two common features, 1) Syntactical or Structural Complexity, and, 2) Lexical Complexity. The former is most commonly coded with regard to the ratio of clauses to the AS-Unit, T-Unit, or C-Unit in the participants’ language production and the latter is measured, for example, by calculating the ratio of lexical words to function words (lexical density).

Once again, however, there is some disagreement among researchers regarding the best unit of analysis to measure Syntactic Complexity. Foster, Tonkin, and Wigglesworth (2000) support the use of an AS-Unit.

The AS-Unit is defined as a “single speaker’s utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (Foster, et al., 2000, p. 365). For example,

A: Which one do you choose?

B: Bold (independent sub-clause)

They also make the point that AS-Units are more appropriate for analyzing spoken language than the other two units, C-Units or T-Units for the reasons that it is

a genuine unit for planning, since many pauses occur in syntactic unit boundaries and it allows for the inclusion of independent sub-clausal units, which are common in speech and syntactic unit offers an acceptable level of reliability, given that syntactic units are easier to identify than intonational and semantic units (p.366).

Norris and Ortega (2009) provide further support for this arguing that “the utterance of AS-Units is considered more appropriate for dialogic oral data, which contain many non-syntactic segments.” (p. 560). Further, Foster et al. (2000, pp. 358-360) identify problems with using C-Units and T-Units. Firstly, since C-Units deal mainly with semantic criteria, there is a question about reliability. Secondly, C-Units which are primarily concerned with “pausing and intonational features” may be problematic because pauses which occur in the L2 oral production are “not necessarily at unit boundaries and it can be difficult to distinguish between pauses that result from message formulation or a lexical search” (p.359). With regard to T-Units, Foster et al., argue that it is not appropriate to deal with a full analysis of spoken discourse using T-Units as they purely focuses on syntactic criteria.

According to Foster et al., another advantage of using AS-Units over C-Units is that AS-Units include the analysis of independent sub-clausal units, which commonly occur in oral production. Independent sub-clausal units may consist of either one or more phrases which can be elaborated by a full clause by referring back to the context of the discourse or situation or a minor utterance defined as “irregular sentences” or “non-sentences”.

1. A: How long have you stayed here?
B: Three months (independent sub-clause)
 2. Oh poor man! (non-sentence)
 3. Thank you very much
 4. Yes
- (Foster et al., p.366)

Furthermore, Foster et al. (2000, p. 370) suggest that the analysis of the highly interactional data should exclude two types of data for the purposes of coherent analysis.

1. One word minor utterances, (e.g., yes, no, okay, right, etc.)
2. Echo responses which are verbatim:

A: I think two years

B: Two years

In short, complexity measures are those that show: 1) Structural or Syntactical Complexity and, 2) Lexical Complexity particularly in the studies of task complexity manipulating planning time (+/- planning time) and the number of elements (+/- few elements). It is argued that AS-Units are more appropriate to analyze oral production data than C-Units or T-Units. For this reason this is the measure of complexity that was used in the current study. At the same time how complexity relates to the performance of the learner from a holistic perspective is less clear. The current research addressed this issue, specifically by using L1 speaker's judgments of learner performance (see 2.7 for a discussion of this).

2.6.2 Accuracy

Accuracy is also defined in different ways by different researchers (Skehan & Foster, 1999; Yuan & Ellis, 2005). Skehan and Foster (1999) define accuracy as:

The ability to avoid error in performance, possibly reflecting higher levels of control in the language as well as a conservative orientation, that is, avoidance of challenging structure that might provoke error (p. 96-97).

In contrast, Yuan and Ellis (2005) specify and define accuracy in terms of error-free clauses and correct verb forms. According to Yuan and Ellis, the former refers to the proportion of clauses that do not contain any error. All errors relating to syntax, morphology, and lexical choice are considered. Lexical errors are defined as errors in lexical form or collocation (e.g., I was waiting you). The latter refers to the proportion of accurately used verbs in terms of tense, aspects, modality, and subject verb agreement. Drawing on Yuan and Ellis' (2005) definition of accuracy, Tavakoli

(2009, p. 488) defines accuracy in terms of error-free clauses, that is “clauses in which [there is] no error with regard to syntax, morphology, native like lexical choice or word order”. For example,

1. I know the man who is sitting next to me (Syntax).
2. My friend is talking with a tourist (Morphology).
3. I am listening to the music (Lexical choice).

Based on these definitions (Foster and Skehan, 1999; Yuan & Ellis, 2005; Tavakoli, 2009) it appears that accuracy is viewed solely from the perspective of what Canale and Swain (1980) earlier labelled ‘grammatical competence’ (i.e., morphology and syntax). Therefore, it seems likely that this measure will simply reflect grammatical accuracy. Even so, like Complexity, Accuracy also has been measured in various ways by different researchers (Robinson, 1995, 2001a; Yuan & Ellis, 2003; Skehan & Foster, 2005; Gilabert, 2005; Michel et al., 2007; Tavakoli, 2009; Ahmadian & Tavakoli, 2010). For example, Robinson (1995) measured accuracy in terms of target-like-use (TLU) of articles (as proposed by Pica, 1984). The use of TLU to measure accuracy of the learner’s language performance appears to be too specific as there are a number of linguistic components other than simply the use of articles. In a later study, Yuan and Ellis (2005) measured accuracy in a more comprehensive manner, using two different calculations, Error-Free Clauses, and Correct verb forms, that is, the proportion of accurately used verbs in terms of tense, aspects, modality, and subject-verb agreement.

Based on this previous research, the current study encapsulated those features of accuracy measures common among a number of studies, namely, Percentage of Error-Free clauses and Ratio of different types of errors (e.g., Percentage of Self-Repairs to Unrepairs, and Self-Repairs to Errors, Error of verb forms, etc). The first of these, Error-Free Clauses, are defined as clauses which do not contain any errors with regard to syntax, morphology, and native like lexical choice or word order (Yuan & Ellis, 2003). It is argued that it best represents the accuracy learner performance in terms of syntax, morphology, and native like lexical choice or word order (Tavakoli, 2009, p. 488). The second measure, Ratio of different types of errors, has also been frequently used. Self-repairs in particular are argued to be most useful because they show learners’ attempts to focus on form (Kormos, 1999). For

instance, Lyster and Ranta (1997, p.57) argue that repairs constitute the process of retrieving target language knowledge learners already have to revise their language production.

However, the issue of what is exactly 'accurate' from a more holistic perspective can be problematic. Therefore, once more the current study used L1 speaker judgments to address this.

2.6.3 Fluency

The term fluency has also been defined in a number of ways by different researchers. Very early in SLA research history, Fillmore (1979) defined fluency as the ability to 1) to fill time with or "disc-jockey fluency"; 2) to talk in coherent, reasoned, and semantically dense sentences; 3) to have appropriate things to say in a wide range of contexts; and, 4) to be creative and imaginative in language use. The definition of fluency by Fillmore (1979) has multiple perspectives including speech rate (i.e., disc-jockey fluency), linguistic features (i.e., coherence and dense semantic constructions) and pragmatic considerations (i.e., appropriateness with the context).

However, in contrast, Lennon (2000) defines fluency according to speech rate.

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).

In addition, Lennon (1990, p. 391-392) also classifies fluency in both a broad sense and a narrow sense. In a broad sense, fluency refers to global oral proficiency in that a fluent speaker has a good command of the second language that is used for a range of purposes, such as employment or for academic reasons (e.g., to study abroad). In a narrow sense, fluency refers to the assessment of one component of oral proficiency such as grading oral examinations in terms of correctness, idiom, relevance, appropriateness, pronunciation, lexical range. This narrower view of fluency appears to be in line with the definition of fluency by Skehan and Foster (1999), that is, "the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems." (p. 96).

Drawing on Lennon (1990), Kormos (2006, p. 155) also classifies fluency in both a broad and narrow sense. First broadly, fluency refers to “global oral proficiency”, such as when a speaker has a generally high level of proficiency in the second language. In a narrower sense, fluency constitutes one component of oral proficiency, which is often used as one of the scores in assessing candidates’ oral language skills in an exam situation.

However, Sajavara (1987) describes fluency differently as “the communicative acceptability of the speech act, or communicative fit” (p.62). Sajavara’s (1987) definition of fluency indicates that fluency closely corresponds to or depends upon the context of where or when speakers have to express their ideas. For example, a speaker is expected to speak more slowly when conversing with children compared to when speaking to adult audiences. Alternatively, a speaker will talk faster when she or he is talking about daily life with a friend. Therefore, a measure of fluency which purely relies on speech rate might be problematic particularly for learners or speakers coming from linguistic cultures where a slow manner of speaking is the custom.

For this reason, Kormos (2006) regards Sajavara’s definition of fluency as problematic in the sense that it is very difficult to operationalize because communicative context will vary in accordance with the situation. Instead Kormos relies on a more discrete perspective suggesting that fluency consists of the number of elements particularly related to speech rate (e.g., words produced per second) and disfluency markers (e.g., repetition and pauses), regardless of the context.

Lennon (1990) also uses discrete measures of fluency, reflecting three main aspects - speech rate, disfluency features (e.g., repetition), and pauses (filled or and unfilled pauses) including 1) words per minute (unpruned); 2) words per minute (after pruning); 3) repetition per T-Unit, self-corrections per T-Unit, filled pauses per T-Unit; 4) percentage of repeated and self-corrected words; 5) total unfilled pause time as percentage of total time of delivery; 6) total filled pause time as percentage of total time of delivery; 7) mean length of speech “runs” between pauses in words; 8) percentage of T-Units followed by pause (filled and unfilled); 9) percentage of total

pause time at all T-Unit boundaries (filled and unfilled); and, 10) mean pause time at T-Unit boundaries (filled and unfilled).

In contrast to this comprehensive list, Bell (2003), who views fluency as a temporal phenomenon, measures fluency using a simplified measure, specifically speech rate, regardless of dysfluency markers and pauses. Other researchers measure fluency simply by calculating the number of complete words produced per second (Riggenbach, 1989; Foster & Skehan, 1996; Mehnert, 1998).

For studies investigating the manipulation of task complexity, fluency has recently been measured using Speech Rate A and Speech Rate B (pruned speech). Speech Rate A refers to the ratio of syllables per minute in unpruned speech. That is, the total number of syllables generated from task performance including, repetition, self-repairs, false starts, etc., divided by the total number of seconds used to perform the task and multiplied by 60. Speech Rate B (pruned speech) refers to the total number of syllables in a dialogue, but it excludes repetitions, self-repairs, repetition, false start, and asides in the L1 divided by the total number of seconds and multiplied by 60 (Yuan & Ellis, 2003; Gilabert, 2005). The measures of Speech Rate A and Speech Rate B are regarded as simplified, but comprehensive measures because they have taken into account the length of pauses and the number of syllables including dysfluency markers (e.g., repetition, false starts, self-repairs, etc.) (Griffiths, 1991; Ellis, 2005). These two types of fluency measures have been used to measure fluency in the following task-based studies: Yuan & Ellis, 2003; Gilabert, 2005; Tavakoli, 2009; Ahmadian & Tavakoli, 2010. The two measures appear to have essentially adopted the fluency measures proposed by Lennon (1990).

Therefore, in this current study, the fluency measures Unpruned Speech Rate A and Pruned Speech Rate B were employed to measure learner language oral production. (Detailed information about these two fluency measures are described in Chapter Three).

2.7 Holistic Assessment of Language Performance: L1 raters' judgments

Assessment of oral language production can be done in two distinct ways, using a set of discrete measures (as described in detail above) or, alternatively, holistically. The holistic model of language assessment makes an overall judgment based on the impression of a speaker's oral production (e.g., Fulcher, 2003). In this way the speaker's abilities to achieve a specific communication purpose is assessed, both linguistically, but also from a pragmatic perspective. This is because, as Kaito and Kaito (1996) argue, testing speaking skills should focus on the learner producing appropriate and meaningful messages, rather than grammatical accuracy. Similarly, Pallotti makes the point that an expression which is grammatically complex and correct is not always accurate from a pragmatic perspective. In fact, the use of discrete assessment (e.g., CAF) which are traditionally used to assess L2 language performance in SLA research, particularly with respect to the manipulation of task complexity, has recently been challenged by a number of authors (e.g., Housen & Kuiken, 2009; Larsen-Freeman, 2009; Skehan, 2009; Norris & Ortega, 2009; Pallotti, 2009). Housen and Kuiken (2009, p. 5), for instance, describe problems of "operationalization" in terms of validity, reliability, and efficiency. Norris and Ortega (2009) describe the redundancy of CAF, particularly the complexity measures. They argue that different complexity measures with respect to subordinate or dependent clauses are "redundant and exactly measure the same things" (p. 560). Larsen-Freeman, (2009, p. 580) makes the point that the use of CAF as measurements of learners' performance on tasks to date is due to "the lack of suitable measures in SLA research". She further states that CAF as the instruments to measure learners' performance are "too blunt". However, Larsen-Freeman (2009) does not suggest an alternative for assessing learners' language performance.

Skehan (2009, p. 510), as one of the proponents of CAF, does make the suggestion that CAF should be supplemented by an examination of lexical use to measure students language performance more effectively. Pallotti (2009) also suggests that there is a need to make a clear distinction between CAF and notions such as interlanguage development and communicative adequacy. Accordingly, Pallotti suggests that communicative adequacy, namely the appropriateness to

communicative goals and situations, “should be considered as a separate measure from CAF or as another measure to assess learner language performance in addition to CAF” (p. 590). One way to do this is through native (L1) speakers’ judgments of L2 language performance. By doing this they might see fluency from a perspective, which House (1996, p. 228) refers to as “pragmatic fluency”, that is, a “dialogic phenomenon that combines both pragmatic appropriateness and smooth continuity in ongoing talk” rather than fluency as “speech rates” on which CAF primarily rely. As such, what CAF measures regard as ‘fluent’ might not necessarily be seen as ‘fluent’ by L1 speakers’ judgments and vice versa.

Overall, it does seem that because of the limitations of CAF, learner performance needs to be evaluated from different perspectives (e.g., L1 speakers’ judgments). Employing L1 speakers as raters to assess L2 oral language performance might provide a complementary analysis of L2 language performance which is not accounted for by CAF measures. This is because L1 speakers’ intuitions of their language are often accurate (Davis, 2003, 2011) because they have “communicative competence”, and can judge grammatical and sociolinguistic competence (Canale and Swain, 1980, p.6). Expressed in another way, they have what Bachman (1990) refers to as “Communicative language ability”, that is,

“The ability to use language communicatively involves both knowledge of competence in the language, and the capacity for implementing, or using this competence in appropriate, contextualized communicative language use” (p. 81).

This suggests that L1 speakers not only have linguistic competence of the language (i.e., they know “what”), but they are also able to perform it in an appropriate way in accordance with the context (i.e., they know “how”). However, it should be noted that although L1 speakers can make accurate judgments, they are not always aware of their linguistic knowledge in a formal sense, nor can they explain how they use their language appropriately or accurately (Cook, 1999).

Davies (2011) points out that one of advantages of using L1 speakers is that “native speakers have intuitions (in terms of acceptability and productiveness) about his language” (p. 303). It is, therefore, reasonable to regard L1 speakers as the “norm” in judging L2 language performance. Studies involving L1 speakers as raters to assess

ESL or EFL oral performance have been numerous. For example, studies have been conducted exploring L1 speakers of different professional backgrounds (Haden, 1991). Studies have also been conducted comparing L1 speakers and L2 speakers judgments (Fayer & Krasinski, 1987; Haden, 1991; Kim, 2009; Zhang & Elder, 2010).

With regard to different L1 language backgrounds, Fayer and Krasinski (1987) investigated oral language performance of EFL Puerto Rican learners of English judged by L1 speakers of English and L1 speakers of Spanish. The study revealed that L1 speakers of English were irritated by particular speech features, especially pronunciation errors and hesitations. Another similar study by Kim (2009) investigated native and non-native teachers' judgments of oral performance of ten Korean learners of English. The study included two groups of raters, each of which consisted of 12 L1 Speakers of (Canadian) English teachers and 12 non-L1 speakers of Korean English teachers. The study revealed that L1 Speakers assess and elaborate L2 performance in more detail than non-L1 speakers in terms of pronunciation, specific grammar use, and accuracy of transferred information. Concerning pronunciation, L1 speakers predominantly commented on a certain feature of pronunciation, such as individual sounds which were incorrectly pronounced (e.g., 'saw' instead of 'show'). Similarly, L1 speaker raters also commented on 'accuracy of transferred', which refers to grammatical features (i.e., the appropriate use of preposition, verb tenses) rather than pragmatic appropriateness. Fluency was also mentioned as one of the assessment criteria of L2 oral performance in terms of "pausing" and "smooth flow of speech".

While L1 speakers in studies of L2 oral performance (Fayer & Krasinski, 1987; Kim, 2009) focused on pronunciation, in particular specific individual sounds (consonant sounds), Jenkins (2002) found broader features of pronunciation often caused communication breakdown because of intelligibility problems, for example consonant sounds, tonic, stress, vowel length, and non-permissible simplification of consonant clusters. Jenkins also identified concerns with pronunciation features, namely the speech sounds.

In their study, Zhang and Elder (2010) compared the holistic scores and the norms applied by 19 L1 speakers of English and 20 L2 speakers of English of 30 Chinese learners of English for oral language performance. Although no significant difference was found between L1 and L2 speaker raters' holistic judgments, the two groups of raters' comments revealed different priorities for a number of various features of the oral proficiency construct. For example, L1 speakers commented more frequently on fluency than did L2 speakers. Conversely, they mentioned 'linguistic resources' less frequently. In addition, they commented on fluency in terms of hesitation and pausing, and rate of speech. This indicates that L1 raters consider not only fluency (in terms of speech rate), but also pragmatic fluency. However, the study did not specifically elaborate the features of pronunciation, that is, whether both raters commented only on certain features (i.e., individual sounds) as found in Kim (2009), or broader features of pronunciation as found in the study by Jenkins (2002). Although L1 speaker norms have been employed to assess L2 oral performance from different perspectives (Fayer & Krasinski, 1987; Haden, 1991; Kim, 2009; Zhang & Elder, 2010), so far, few studies have attempted to explore the degree of fit between learners' L2 performance assessed by L1 speakers of English and traditional CAF measures. This was one of the aims of this study to do so.

2.8 Summary

This literature review has described tasks, TBLT and task-based research. Descriptions were provided of task-based studies undertaken from an information-processing perspective in which the manipulation of task complexity (cognitive factors) was done to test Robinson's (2001a) Cognition Hypothesis. However, there are queries with respect to these studies. Firstly, the studies have mostly examined the manipulation of task complexity under one dimension, either the resource-directing or the resource-depleting dimension, rather than simultaneously investigating the manipulation of the two dimensions. Secondly, studies of tasks from an information-processing perspective have mainly been conducted using monologic tasks. Thirdly, the investigation of learners' perceptions of the tasks has mainly been conducted on the basis of a quantitative approach using Robinson's (2001a) closed-ended questions, rather than using qualitative methods and open

ended questions in particular. Finally, learners' language performance on task manipulation has been commonly measured using CAF, although the use of CAF has recently been challenged within SLA research. The use of holistic assessment, and specifically L1 speaker judgments have been suggested in the literature as an alternative.

On this basis, this research sought to answer the following three research questions:

1. To what extent do dialogic tasks manipulated simultaneously along the resource-directing and the resource-depleting dimensions (i.e., planning time and the number of elements) affect the complexity, accuracy, and fluency (CAF) of learner production?
2. Is there a relationship between students' perception of task difficulty and the four levels of dialogic task difficulty manipulated simultaneously within planning time (+/– *planning time*) and the number of elements (+/– *few elements*)?
3. What is the fit between CAF measures of learners' language production and the judgment of L1 raters?

Chapter Three

METHODOLOGY

3.1 Overview

In this chapter the methodology used in the current study is described. First, the research design and data collection procedures, consisting of the pilot testing and implementation stages, are outlined. This is followed by a description of the participants in this study and the materials used. Finally, the data analyses are discussed.

3.2 Design of the Research

Over the last few decades there has been a trend in research from simply using either a quantitative or qualitative approach to combining both for mixed model studies (Tashakkori & Tedlie, 1998) or mixed methods (Creswell & Plano, 2007; Creswell, 2008). Tashakkori & Tedlie, (1998, p. 19) define mixed model studies as “studies that are products of pragmatists paradigm and that combine qualitative and quantitative within different phases of the research process”. As indicated in Chapter One, the objectives of the current study are to investigate dialogic tasks and their effect on Complexity, Accuracy, and Fluency (CAF); participants’ perceptions of these tasks; and the fit between CAF measures and L1 raters’ judgments of learner proficiency, based on their speech production as they performed the tasks. To do this, the study employed both quantitative methods (i.e., the effect of task difficulty on CAF) and qualitative data (i.e., the Indonesian participants’ perceptions and L1 raters’ assessments). Therefore, this study is mixed methods in design (Creswell & Plano, 2007; Creswell, 2008). It also adopted Creswell’s (2008) model of triangulation in the sense that two different sets of data from “quantitative and qualitative analyses were compared to determine whether or not the two databases supported to each other” (p. 558).

A mixed methods design was used because as Creswell and Plano (2007) indicate neither quantitative approaches, nor qualitative approaches are by themselves, sufficient to answer the complexity of the research problems. Further, the combination of both quantitative and qualitative methods provides a better understanding of the research problems and questions than employing just one approach (Creswell, 2005, 2008; Creswell & Plano, 2007).

As outlined above, triangulation was also undertaken in the current study and this was made possible by using a mixed methods design in which quantitative and qualitative data were merged by data transformation and discussion (Creswell & Plano Clark, 2007). According to Tashakkori and Tedlie (1998), data transformation is a process of ‘quantizing’ qualitative data/ information or “qualitizing” quantitative data/information, although it should be noted that transforming qualitative data into quantitative data is easier than vice versa (Creswell & Plano, 2007). Onwuegbuzie and Teddlie (2003) make the point that the purpose of data transformation is to compare two types of data for further analyses. This was done in the current study when quantitative data were compared with qualitative data and where some of the qualitative data was quantified (namely when the participants’ perceptions were tabulated into percentages and compared).

With respect to the strategy of merging through discussion, as suggested by Creswell and Plano (2007), this can be done, for example, by first presenting a quantitative result and then following it up with a description using qualitative quotes; or the qualitative results are first presented and then followed by the quantitative data. In this study this was done in two ways. Firstly, the four levels of dialogic task difficulty as determined by the task conditions (planning and complexity) were compared with the participants’ perceptions of task difficulty (learner factors). Secondly the quantitative data of oral production in terms of CAF were compared with L1 raters’ judgments. This was done to investigate the degree of fit between the oral production of the Indonesian participants as measured by CAF and L1 raters’ judgments, the data for which consisted of rating scales (quantitative) and written comments (qualitative).

3.3 Data Collection Procedures

The data collection for this study was undertaken in two main stages 1) the preparation stage, and, 2) the implementation stage.

1) Preparation stage

This stage involved designing three types of research instruments, trials of these instruments, analysis of the trials and revision of the instrumentation (see Figure 4). This stage was undertaken to ensure that the instruments (i.e., four types of dialogic tasks the interview protocols, and L1 raters' rating scales) used in the research were valid and reliable.

Each of these instruments was designed to help answer a specific Research Question, namely:

Tasks – Research Question 1

Interview protocols – Research Question 2

L1 raters' rating scales and written comments – Research Question 3

The preparation stage of this study was done in four steps (as shown in Figure 4 below). This consisted of: i) an initial development of the instruments (i.e., tasks, interview protocols, and L1 raters's rating scales); ii) trials of tasks, interview protocols (audiotaped), and L1 raters's rating scales (Note: Each type of instrument was trialed three times); iii) analysis of the trials of the instruments; and, iv) revisions of all the instruments. Descriptions of the instruments are provided in detail in section 3.4 - Materials and Equipment.

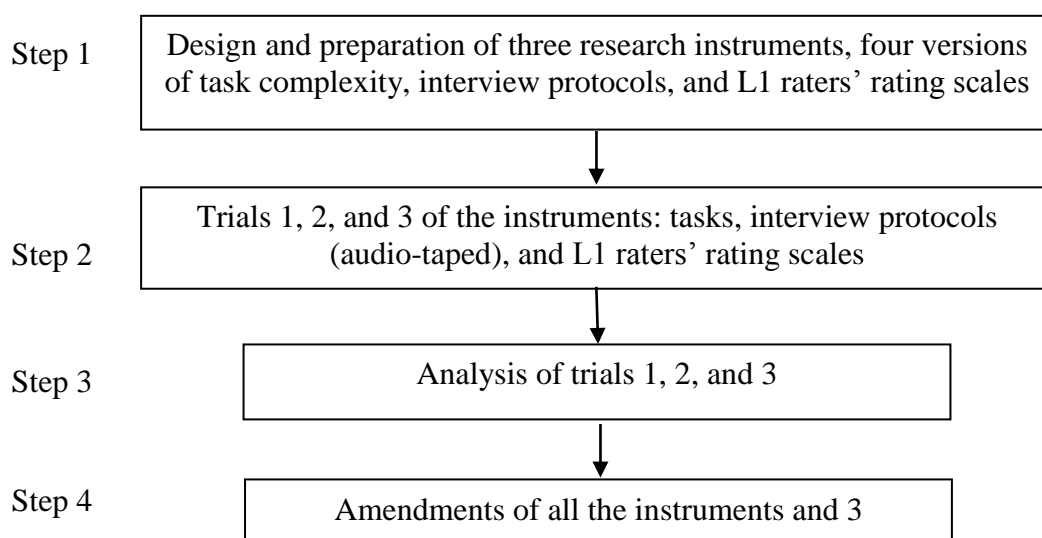


Figure 4: Diagram of preparation stage

2) Implementation stage.

The implementation stage consisted of three phases.

Phase 1 included the collection of both quantitative and qualitative data related to learners performing the dialogic tasks. To do this, the participants were first randomly paired. The researcher was in attendance as each pair performed the tasks. The data was then collected in four rounds according to the sequential number of the task.

As the pairs performed the tasks, they were audio recorded using an Olympus digital recorder. These recordings were then transcribed and coded for CAF (quantitative data). This was followed by interviews with the participants regarding their perceptions of the tasks. The interviews were conducted in English, however, Indonesian (Bahasa Indonesia) was also used as necessary for clarification.

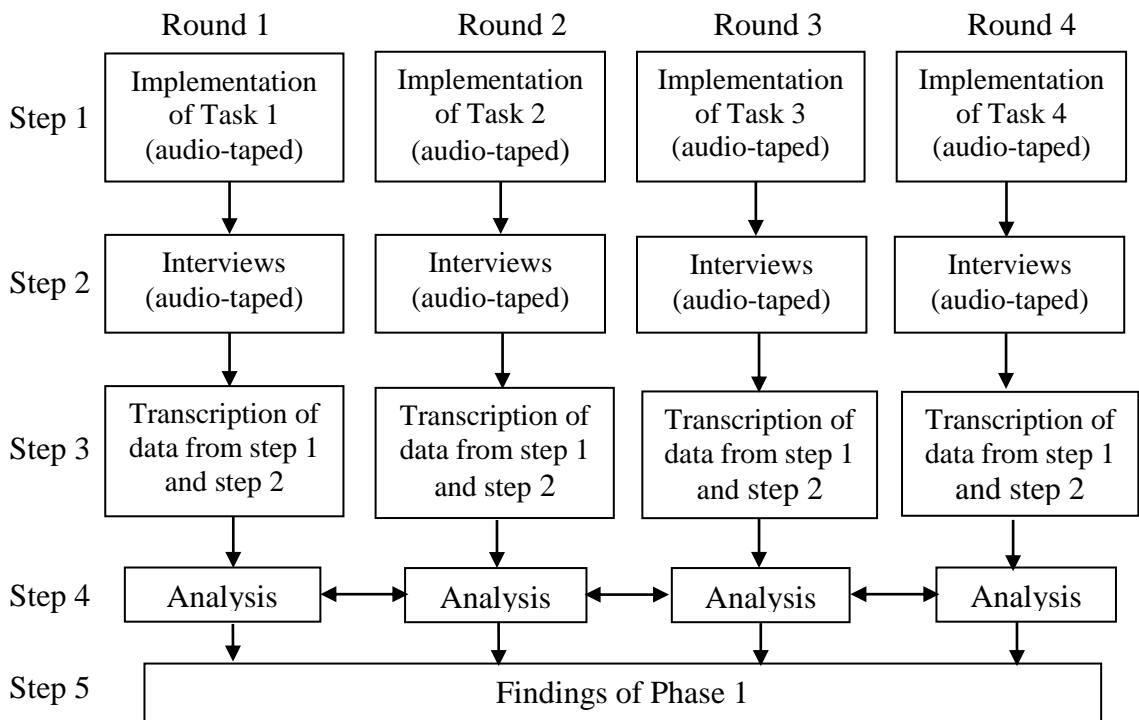
Phase 2 involved the exploration of the L1 raters' judgments of the participants' speech production that occurred during the task implementation. The three L1 raters were asked to rate and give written comments on the data collected during Phase 1, namely the Indonesian participants' speech production. A rating scale was used to measure their perceptions of the overall proficiency of each of the participants and comments about their judgments were also elicited. Accordingly, the L1 raters with

non-linguistic backgrounds were expected to judge the Indonesian participants' oral production in two ways.

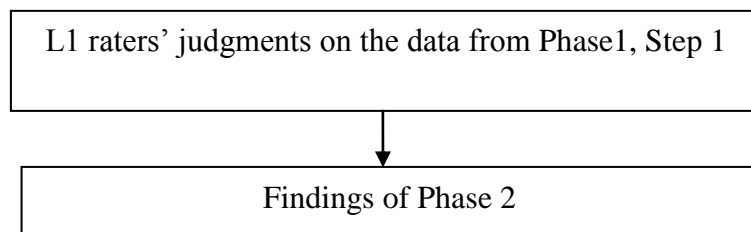
Phase 3 was the comparison of oral production as measured by CAF and according to L1 raters' judgments. The comparison of the two sorts of data allowed for an examination of the degree of fit between these two methods.

The implementation phase of the research is summarized in shown in Figure 5 below.

Phase 1



Phase 2



Phase 3 Analysis Stage

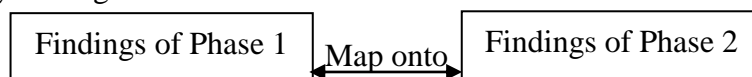


Figure 5: Three phases of the implementation stage

3.4 Participants

There were two groups of participants in this study, Indonesian EFL learners and Australian English L1 raters. The former were those who participated by performing the four tasks and undertaking the post task interviews (Phase 1) and the latter were those who took part in the L1 raters' judgments of the oral production by the Indonesian participants (Phase 2).

The participants of Phase 1 were first year students enrolled in an English Study Program (ESP) at the University of Lampung (UNILA), Indonesia. Based on the selection criteria for this program, administered through the national higher education entrance test (SNM-PTN), their proficiency levels were deemed to be relatively similar. However, to ensure this was the case, the English proficiency test (EPT) was also administered to 60 potential participants before the research was conducted. Based on the test results, four participants were excluded from this study due to them having markedly different EPT scores compared to the remainder of the cohort. While one participant had a lower band score, the other three had much higher scores. In addition, two participants were excluded as they participated in the trials of the instruments, and the other two were absent during the data collection without prior notification. Thus, there were 52 students participating in this research. This number of participants met the criteria for using statistical analysis, which according to Cohen, Manion, and Morrison (2007) should be more than 30 participants. Of the 52 participants, 14 were male and 38 were female and all were aged between 18 and 20 years. They had all formally studied English at high school for six years and for approximately six months in the ESP, at UNILA.

With respect to the participants of Phase 2, there were three L1 raters of Standard Australian English (SAE) who participated as raters. One was male and two were females and they were aged between 50 and 55 years. Only three L1 raters were used in the current study because of the commitment required: Specifically the four data sets took approximately 20 hours in total to rate. However, the long period of time involved allowed repeated and quality listening. The SAE L1 raters were chosen especially to be from non-linguistic backgrounds because one aim of the research is to compare CAF measurements of proficiency with non-specialist L1 raters'

intuitions concerning proficiency. This was done rather than using L1 raters who have had formal training in linguistics because they might be expected to judge speech production relying more on linguistic considerations (Haden, 1991).

3.5 Materials and Equipment

As outlined above, a key aspect of this study was the use of dialogic tasks of various levels of difficulty. Those developed in the preparation stage and then used in the implementation stage (Phase1, Stage 1) are described below. In addition, information about the instruments and the equipment used for the interview protocols (Phase 1, Stage 2) and L1 judgments (Stage 2) are presented. Finally, it should be noted, each instrument was trialed and further developed in an iterative process.

3.5.1 Tasks

Four difficulty levels of dialogic tasks were used in the current study. They were designed in such a way that the factors of planning time (+/- *planning time*) and the number of elements (+/- *few elements*) were manipulated simultaneously. These factors were chosen based on information from the literature, namely that providing planning time can help learners focus on form (Ellis, 2003), that previous studies showed positive effects of planning time, particularly ten-minute planning time on learners' language production in terms of Complexity, Accuracy and Fluency (CAF) (Foster & Skehan, 1996; Mehnert, 1998; Yuan & Ellis, 2003; Ellis, 2004; Gilabert, 2005; Philp, et al., 2006) and, that tasks which are manipulated according to the number of elements (+/- *few elements*) are expected to be more comprehensive in the sense that learners might inevitably include the other two factors of resource-directing dimensions viz. giving reasons (+/- *no reasoning demands*) and using present or past activities (+/- *here and now*) while performing the tasks. Moreover, studies investigating the manipulation of numbers of elements suggest that these aspects have enabled learners to improve their language performance in terms of CAF (Robinson, 2001a; Gilabert, 2005; Kuiken & Vedder, 2007; Mitchel, et al., 2007).

The dialogic tasks used in the current study were based on those tasks used by Michel et al., (2007). This is because these tasks were not only relatively recently developed (i.e., less than five years prior to the commencement of the data collection), they were also specifically designed as dialogic (two-way interactive) tasks – a focus of the current study. Further, they were designed to be either simple or complex and based on the number of elements (+/– few elements) as was the intention of this research. Therefore, reliability of the instruments was made.

In the current study the tasks were developed to provide four levels of task difficulty, two simple and two complex tasks. They were manipulated simultaneously within two dimensions, the resource-directing and the resource-depleting dimensions, namely, the factors of planning time (+/– planning time) and the number of elements (+/– few elements). The two simple tasks had few elements and were undertaken with or without planning time (+/– planning time). According to Robinson (2003, 2005), the former are called Low Performative and Low Developmental Complexity while the latter are High Performative and Low Developmental Complexity. The complex tasks were undertaken with planning time (+ planning time, – few elements) or without planning time (– planning, – few elements). The former are called Low Performative and High Developmental Complexity and the latter are High Performative and High Developmental Complexity (Robinson, 2003, 2005). The organisation of each version of the dialogic task difficulty is set out in Figure 6 and details about each task are provided below.

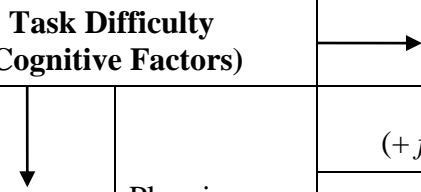
Task Difficulty (Cognitive Factors)		 The resource-directing dimension (+/– <i>few elements</i>)	
The resource-depleting dimension (+/– <i>planning time</i>)	Planning time (+ <i>planning</i>)	Simple (+ <i>few Elements</i>)	Complex (– <i>few elements</i>)
	No planning Time (– <i>planning</i>)	Task 1 The planned simple task (+ <i>planning</i> , + <i>few elements</i>)	Task 3 The planned complex task (+ <i>planning</i> , – <i>few elements</i>)
		Task 2 The unplanned simple Task (– <i>planning</i> , + <i>few elements</i>)	Task 4 The unplanned complex Task (– <i>planning</i> , – <i>few elements</i>)

Figure 6: Four levels of (dialogic) task difficulty developed in this study

The topics of the tasks - Blackberry mobile phones and Houses for Rent - were selected in accordance with the participants' backgrounds. Mobile phone devices, especially Blackberry brands, and Houses for Rent were chosen because both of the topics reflect the life experiences of the participants. In addition, the topics of the tasks allowed for easy manipulation of the planned and unplanned conditions. This was done to minimize the effect of repetition due to familiarity with the same topic rather than the manipulation of planning time (+/- *planning time*) and the number of elements (+/- *few elements*).

Task 1 - The Planned Simple Task

Task 1 (+ *planning time*, + *few elements*) comprised two different pictures of Blackberry mobile phone devices and the participants were given ten minutes of planning time to plan what they were going to say before they performed the task (see Appendix 1). The participants were also provided with instructions to perform the task, including explaining that they should give their partners information or specifications about Blackberries, such as price, size, display, ring tones, memory, colors, and features, and so on. As a familiar topic, it was expected that the task would stimulate the learners to speak. According to the Cognition Hypothesis, Task 1 would be predicted to be the easiest task for the participants to perform because the task is provided with planning time and it consists of few elements (i.e., only two mobile phones).

Task 2 - The Unplanned Simple Task

This was also designed as a simple task, with few elements, but for this task no planning time was provided. Thus, Task 2 is an unplanned simple task. Unlike Task 1, the topic of Task 2 was Houses for Rent, that is, two different pictures of houses were provided for the participants to describe and then select (see Appendix 2). It was chosen as a familiar, but different topic to Task 1. The written instructions about performing the task included telling the participants to provide their partners with information about the houses (e.g., price, location, facilities), however, they were only allowed approximately 1 minute to read these.

Based on the Cognition Hypothesis, because of the lack of planning time, it was predicted that Task 2 would be more difficult than Task 1. Although Task 2 differs in topic from Task 1, both of the tasks could be regarded as essentially the same as they have two similar characteristics. Firstly, each of the tasks comprises the same number of elements (two pictures) and each includes specifications about the topics. Secondly, the topics of both tasks (Blackberry mobile phones and Houses for Rent) are familiar to the participants.

Task 3 - The Planned Complex Task

Task 3 was a complex task as it included many elements. However, ten minutes of planning time prior to performing the tasks was provided. Thus, Task 3 is a planned, complex task (+ planning time, – few elements). The task consisted of pictures of six different types of Blackberries, each with different features (e.g., prices, colors, weights) for the participants to discuss and choose (see Appendix 3).

Therefore, although similar in some ways (i.e., topic and goal), Task 3 differs from Task 1 in terms of the degree of difficulty. This was done to prevent the participants from merely memorizing the same types of Blackberries used in Task 1. Based on the Cognition Hypothesis, Task 3 should be more cognitively demanding than either Task 1 or 2. It is more complex within the resource-directing dimension (– *few elements*), but simultaneously simpler within the resource-dispersing dimension (+ *planning time*).

Task 4 - The Unplanned Complex Task

Task 4 was made complex as it had many elements (– *few elements*), but the participants were not given ten minute of planning time prior to performing the task. Thus, Task 4 is an unplanned complex task (– *planning time*, – *few elements*). According to the Cognition Hypothesis, Task 4 would be the most difficult task for the participants to perform among the four levels of tasks, as it is complex within both dimensions, that is, resource-directing and resource-dispersing. Again the topic of Task 4 was similar to that of Task 2, that is, Houses for Rent, but the two tasks differed in terms of the number and the models of the houses – once more to

minimize the effect of memorizing the same houses included in Task 2. Task 4 included six types of houses and the participants were only provided with instructions of how to perform the task (e.g., specifications, price, facilities, location, etc.), but given no planning time (See Appendix 4).

3.5.2 Interview Protocols

As described in Chapter Two, the interview protocols used in this study were developed based on Robinson's (2001) five closed-ended questions with each taking the form of a nine-point Likert scale. The questions consisted of five categories: levels of difficulty, degree of stress, confidence, interest in task content, and motivation. However, Robinson's (2001a) model of rating scales is quantitative in nature and so it does not allow for the type of qualitative analysis that enables researchers to explore in depth participants' perceptions of the tasks. Therefore, in the current study the interview protocols that were developed were designed to elicit such data. As such Robinson's closed-ended questions were modified into semi-ended and also open-ended questions for the purpose of qualitative interviews because, as Creswell (2009, p. 225) argues "open-ended questions will best enable the participants to express their experiences unconstrained by any perspectives of researchers or past research findings". That is, the questions were converted to *What-questions* followed up by *Why-questions*. In addition, since the tasks that were carried out were dialogic, the interview was also extended with additional questions adapted from Kim (2009) (See Appendix 5 for a copy of the interview protocols). Each interview was audio recorded using the equipment as described in 3.2 above.

3.5.3 The L1 Rating Scale

To enable the L1 raters to make judgments of the participants' language proficiency based on their performance of the four tasks a quantitative instrument was developed. It took the form of a ten-point rating scale, ranging from 'Very poor' to 'Very good' and was followed up with open-ended questions (see Appendixes 6 to 9). A ten-point Likert scale was chosen in the current study for two reasons. Firstly, even with clear instruction about how to make these judgments, those judging are

often unfocussed and, further, making such judgments of speakers of other languages (whether L1 or EFL/ESL) are based on subtle and difficult to capture distinctions (Davies, 2003, 2011). Thus a scale was developed for this research to allow the kind of nuanced judgments that would not have been possible, without a scale or by using one with fewer options (e.g., a five- or a seven-point scale). Secondly, a ten point–Likert scale has been deemed the most useful in previous research (Preston & Colman, 2000) and based on the trial of the instrument (see 3.5.4 below).

3.5.4 Trialing the Instruments

Once the initial instruments had been developed, trials were conducted. This was done to establish both the strengths and the weaknesses of all the instrumentation. Based on this trialing, amendments were made. As indicated, this was an iterative process and each type of instrument was trialed three times: the first two trials of task difficulty and interview protocols were conducted with Indonesian EFL learners of English studying in Perth, Western Australia and the third trial was administered in Indonesia with some of the first year students enrolled in the English Study Program at the University of Lampung (UNILA). The trials of L1 rater's rating scales were all conducted three times with L1 raters in Australia.

Tasks

Trials of these tasks were deemed necessary to ensure that they were appropriate for the L2 level of the participants (Ellis, 2009, p. 241). The first trial was conducted with two adult Indonesian learners of English aged between 27 and 30 years old.

With Task 1, that is, the planned simple task they were given ten minutes and were provided with a piece of paper and pens to take notes of what they wanted to say before performing the task. However, neither used the planning time nor took notes of what they planned to talk about. They just glanced at the pictures and read the instructions for a few minutes instead. They did not discuss nor compare individual specifications or information provided on the task, but rather discussed general information of the two types of the mobile phones. They completed Task 1 in approximately six minutes.

For Task 2 the trial participants were not given ten minutes of planning time, instead they were told to read the instructions for approximately one minute. Task 2 was also completed in approximately six minutes. They were silent for a few minutes before they performed the task. They seemed to be thinking about *what* and *how* to start the conversation.

Drawing on the weaknesses found in the first trial, amendments were made. Specifically, with respect to planning time, the participants were given some direction about how to plan, for example, they were told to take notes about the strengths and weaknesses of the phones provided in the task. This proved to be a useful modification.

The second trial was undertaken with two adult Indonesian students aged between 35 and 40 years old. All four tasks were trialed. Using the new instructions in the second trial for tasks one and three, the participants did plan what they wanted to discuss. They took notes for approximately ten minutes. They performed the tasks in the second trial for around ten minutes. Although the participants did not compare every single specification of the pictures provided in the tasks, they discussed more features of the pictures compared to those in the first trial. However, it did seem that they found it difficult to express all their ideas appropriately in English.

Procedurally in the first and second trials, the roles of speakers performing the tasks did not alternate. That is, Speakers A and B had the same roles in all four levels of tasks. In the third trial this was changed and the roles of the speakers were alternated.

The third trial was conducted in Indonesia with two Indonesian female participants studying in the English Study Program at the University of Lampung. They were between 18 to 20 years of age and were drawn from the pool of potential participants.

It was also clear that in the second trial there was a need to provide even more detail in the instructions about how to perform the tasks so that the participants could clearly understand what they needed to do. This was done and it was found by the third trial this enabled them to discuss the topic for a longer period of time.

Interview Protocols

The trial of the interview protocols was conducted to make sure the questions were effective for exploring the participants' perceptions of the tasks and their difficulty. The first trial of the interview protocols was carried out shortly after of the first trial the tasks.

It was found that the participants did not seem to be familiar with the term "task". They commented more on the pictures rather than on their perceptions of the planned and unplanned simple tasks. Thus, in the second trial the term "task" (e.g., a language learning activity) was first defined and then used in the interviews.

Furthermore, it was found that the participants had difficulty understanding a few questions. For example, question number 5, "How well could you complete the task?" was deemed to be too difficult for the participants to answer, therefore, the question was amended to include a follow up question (i.e., "Did you perform the task well or not well?"). The trial participants also appeared to find confusing the follow-up question to "How did you feel while performing the task, for example, did you feel relaxed or frustrated?", namely "What do you think made you feel like this?" This was amended for the next trial.

In the second trial, the participants were not found to have any further problems with the questions. Since all four levels of task difficulty were trialed in the second iteration, the interviews were carried out about all four levels of tasks. However, it was found that the second speaker often gave the same response as the first. This was dealt with in the third trial as described below.

The third trial of interview protocols was also conducted in Indonesia with the two female Indonesian participants who performed the tasks in the third trial. To address the similarity of responses, the participants were alternated from that in the task (e.g., Speaker A was asked the interview questions second, Speaker B asked first), to minimize the effect of imitation. The result of the third trial with interview protocols suggests that the questions could be easily understood by the participants, and they

were useful for exploring the participants' perceptions of task difficulty. Therefore, this was the form of the instrument used during the implementation stage.

L1 Rating Scale

The trials of the L1 rating scale were also carried out three times. In the first trial, the rating scale used a six-point Likert Scale, ranging from 'Worst' to 'Best'. The trial was conducted with one L1 rater of English. The speaker was asked to rate and comment on each the Indonesian learners of English in terms of how good he or she was at speaking.

This six-point Likert Scale did not seem to allow the rater to make nuanced judgments. Moreover, the use of the terms 'Worst' and 'Best' in the first L1 rating scale seemed to be less appropriate as the terms are related to a mental state judgment rather than to language performance. Thus, the first L1 raters rating scale was amended for the second trial, both in terms of the number of the number of options and the descriptors used. Specifically, in the second trial, the L1 rating scale used twelve points and the terms 'Worst' and 'Best' were replaced with 'very poor' and 'very good'.

The second trial of L1 raters' rating scale was also conducted with an L1 rater who was a speaker of Standard Australian English. The twelve-point rating scale appeared to be too broad for assessment purposes. In addition, it seemed that the instructions were not clear enough.

These weaknesses were addressed in the third trial. Firstly, a ten-point Likert scale was used and with respect to the instructions, these were amended to improve clarity.

The third trial was undertaken with one L1 rater of English. It was clear that the instrument allowed the L1 rater to better discriminate between the participants' oral production. It was also found that the instructions were appropriate.

3.6 Data Analysis

Following the data collection, the three sources of data were analysed according to: i) participants' oral production about the four tasks; ii) the participants' interviews; and, iii) L1 raters' judgment on the participants' oral language production. The analysis of each was conducted in the manner as described below.

3.6.1 Analysis of Participants' Oral Production about Tasks

The participants' oral production was analyzed in terms of CAF. To do this, first the recordings of the participants performing the tasks were transcribed, and coded for the various CAF measures.

These CAF measures were adapted from those used in the study by Michel et al. (2007). It entailed the use of multiple aspects of CAF, including:

- Complexity (Both syntactic and lexical complexity were calculated):
 - Syntactic Complexity was calculated manually by determining the AS-Units (the analysis of speech units). This was used instead of T or C-Units because the interactional nature of the data meant that it consisted of many non-syntactic segments (Foster, et al., 2000; Norris & Ortega, 2009).
 - Lexical Complexity, as measured by the Percentage of Lexical Words to a Total Number of Words, done using the Conversation Analysis Mode of CHILDES (McWhiney, 2000) (also see Table 2 below adapted from Gilabert, 2005).
 - Guiraud's Index of Lexical Richness, also using CHILDES.

Table 2: Calculation of Lexical Words

No	Lexical Words	Examples
1	full verbs, nouns, adjective, adverbs ending in <i>ly</i>	buy, houses, good, carefully
2	the verbs <i>have, do, be</i> except when used as auxiliaries	I <i>have</i> much money.
3	wrongly conjugated verbs	Buyed
4	words that have problems with number	man, men
5	interjections	hi, hello, goodbye
6	hyphenated words and contractions	I'm, I'd
7	conjugated forms of verbs count as different types	do and did
8	phrasal verbs	to get up
9	in preposition verbs	interested in

- Accuracy, also calculated manually by determining Percentage of Error-Free Clauses, Percentage of Self-Repairs, and Ratio of Repaired Errors to Unrepaired Errors as shown in Table 3Table 3 below.

Table 3: Measures of Accuracy

Accuracy Measures
<p>1. Percentage of Error-Free AS-Units:</p> $\frac{\text{Number of Error-Free AS-units}}{\text{Number of AS-Units}} \times 100$
<p>2. Percentage of Self-Repairs:</p> $\frac{\text{Number of Self-Repairs}}{\text{Number of Errors}} \times 100$
<p>3. Ratio of repaired errors to unrepaired errors:</p> $\frac{\text{Number of Repaired Errors}}{\text{Number of Unrepaired Errors}} \times 100$

- Fluency was again calculated manually, ascertaining the Unpruned Speech Rate A and Pruned Speech Rate B (Yuan & Ellis, 2003; Gilabert, 2005). (See Table 4 adapted from Gilabert, 2005).

To calculate Speech Rate A, the number of syllables² used per minute was determined, with the following rules applied.

1. *Ing* forms such as, doing, saying, etc., counted as two syllables.
2. The constructions such as, isn't, doesn't, didn't, were calculated as two syllables.
3. Syllables in Indonesian words were counted (Speech Rate A).
4. Epenthesis (insertion of sounds in the middle of words) does not count as a syllable, e.g., speak /səpi:k/, instead of /spi:k/.
5. Past /ed/ form was not regarded as a syllable (e.g., looked). But past /ed/ was calculated as a syllable for the verbs ending with *t* or *d* (e.g., “wanted”, “landed”), each counted as two syllables.

- Speech Rate B was also calculated in a similar way to Speech Rate A, but syllables which appeared as repetitions, self-corrections, false starts, and in Indonesian or local words were excluded as shown in the example below.

Repetition: She she chose Blackberry onyx white.

Repair : I have not much don't have much money.

False start: The two the girl goes to the university every day.

Indonesian words: I have to *apa memilih* /what is “choose” in English?/

But the following examples were not calculated as repetition or self-corrections.

I want to buy a Blackberry a Blackberry Onyx.

² In the current study, a syllable is taken to refer to any “syllable types” of English as elaborated by McKay (2004). These syllable types include a single vowel (V), and vowel consonant clusters, such as, VC, CV, CVC, CVCC, CCVCC, and CCCVCCC.

I think that she your girlfriend loves Blackberry.

Table 4: Measures of Fluency

Fluency Measures	
1. Speech Rate A (unpruned speech):	
$\frac{\text{Number of Syllables}}{\text{Total number of seconds}} \times 60$	
2. Speech Rate B (pruned speech):	
$\frac{\text{Number of Syllables}}{\text{Total number of seconds}} \times 60$	

Once the CAF measures were calculated the results for the four levels of task difficulty (+/- *planning times*) and (+/- *few elements*) were compared using the statistical formula of Repeated Measures Analyses of Variance (ANOVA). This was done to find out whether or not the four levels of task difficulty were significantly different in terms of CAF. The comparisons essentially included six models, but they were merged into three groups according to similar characteristics of the tasks as shown in Table 5 below.

Table 5: Comparisons of four levels of task difficulty

No	Comparisons	Tasks
1	Planning and No planning (+ <i>planning</i> and – <i>planning time</i>) in both simple and complex conditions (+ <i>few elements</i> and – <i>few elements</i>)	Tasks 1 and 2
		Tasks 3 and 4
2	Simple and Complex (+ <i>few elements</i> and – <i>few elements</i>) in both planned and unplanned conditions (+ <i>planning</i> and <i>planning time</i>)	Tasks 1 and 3
		Tasks 2 and 4
3	Planned simple and Unplanned complex, and Unplanned simple and Planned Complex	Tasks 1 and 4
		Tasks 2 and 3

3.6.2 Analysis of participants' post-task interviews

To undertake the analysis of this data, first it was transcribed and then organized systematically in preparation for analysis. Next all the data was read through and coded in detail. From this categories or themes emerged, which were then interpreted for presentation in the findings.

To undertake the coding a binary system was used in which the participants who had opposite responses for each category were designated either (+) or (-). Both plus (+) and minus (-) codes were then accompanied by a number referring to the order of the questions in the interviews. For example, the Plus (+) code was generated from question 1 and was coded by "1+". A minus (-) response generated from question 1 would be then coded by "1-", etc. (See Appendix 10 for an example of this).

The participants' responses were coded manually with reference to Descriptive and In-Vivo Codes (Miles & Huberman, 1994; Saldana, 2009). According to Saldana (2009) the first term refers to the summary of the primary topic of the excerpt, while the latter means a direct quotation taken from what the participant says (pp. 3-4). Drawing on these procedures, the coding process of this study was dealt with as shown in the example below.

It is *easy* I think because we *just compare between Blackberry onyx white and Blackberry bold* I think and we just give our argument why we choose the only of two (Task 1).

The word 'easy' is coded (1+) and the following responses to 1+, "just compare between Blackberry onyx white and Blackberry bold", which mean comparing only two models of Blackberry hand phones, were summarized by a Descriptive Code as a 'simple task' (ST).

The same procedures were applied to minus (-) responses. As shown in the excerpt below.

I think the task *difficult* because we *must choose many blackberries* and it made me confused (Task 3).

The word “difficult” is coded by (1–), while the response following the minus (1–) code, “must choose many Blackberries”, was coded as a complex task (CT).

These data were then tabulated as a percentage agreement summary of all the participants’ perceptions of task difficulty, which is presented in the results (see 5.2 in Chapter Five). This was done to address Research Question 2.

3.6.3 Analysis of L1 raters’ Assessments

The data generated from L1 raters’ assessments were in the form of scores from rating scales and the written comments on the oral production. Because of the use of a broader scale (a ten-point Likert scale) there were some disagreements in ratings. When this occurred, a certain amount of latitude was needed. Therefore, in this research reasonable divergence in ratings was accepted and the data of rating scales were analysed on the basis of the as following criteria:

- (i) when three ratings were the same (e.g., 6 6 6);
- (ii) the two ratings were the same and the third is within one or two-point scale, (e.g., 5 5 6, or 5 5 7);
- (iii) the three ratings were within one-point Likert scale (e.g., 5 6 7); and
- (iv) the two ratings were within one-point Likert scale and the third rating was within two-point Likert scale (e.g., 4 5 7).

However, when data did not meet these criteria, they were rejected or excluded from the analysis (see Appendix 11). However, this analysis was also informed by the comments provided by the three raters and was analyzed using the binary system as described above.

Next the three L1 rating scales were compared with the CAF results. These findings were also reported. Specifically, the patterns of L1 rating scales and written comments were compared with those of CAF for an examination of the degree of fit between these two methods (see sections: 6.1.1, 6.1.2, and 6.1.3, in Chapter Six). This was to address Research Question 3.

3.7 Summary

This study used mixed methodology: collecting and analyzing quantitative and qualitative data. Quantitative (i.e., oral production in terms of CAF) and qualitative data (i.e., participants' perceptions of the four levels of tasks) were collected from the 52 Indonesian EFL learners of English, while other data (i.e., rating scales and written comments) were collected from three L1 raters of Standard Australian English. The data of this research were analyzed through three steps. Firstly, an analysis of the CAF data generated from the four tasks, next, the participants' perceptions of these tasks and, finally, the CAF results were compared with the L1 raters of English judgments.

Chapter Four

FINDINGS ABOUT TASK DIFFICULTY

4.1 Overview

This Chapter addresses Research Question 1 and reports on the findings from the data on Complexity, Accuracy, and Fluency (CAF) generated by the four levels of dialogic task difficulty manipulated simultaneously within the planning time (*+/- planning time*) and the number of elements (*+/- few elements*). It describes the CAF results when the four tasks are compared.

4.2 CAF and the Four Levels of Task Difficulty

This section reports the findings of 52 Indonesian participants' performance in terms of CAF resulting from their performance of the four tasks.

As can be seen in Table 6 the performance of the four tasks resulted in different mean scores on the various CAF measures.

With regard to the measures of **Complexity**, the unplanned simple task (Task 2) had highest mean across all scores (i.e., Syntactic Complexity, Percentage of Lexical Words, and Guiraud's Index of Lexical Richness) compared to the other three versions of the tasks (i.e., Tasks 1, 3, and 4). The planned simple task (Task 1) had the second highest mean of Syntactic Complexity and Percentage of Lexical Words, followed by the unplanned and planned complex tasks (Tasks 4 and 3).

Table 6: Means of CAF for the four levels of Tasks

Task/Measure	Task 1 (+ planning time/+ few elements)	Task 2 (– planning time/+ few elements)	Task 3 (+ planning time/–few elements)	Task 4 (– planning time/(– few elements)
Complexity				
Syntactic: AS-Units	1.54	1.65	1.41	1.42
<i>Lexical:</i> % of Lexical Words to a Total Number of Words	18.84	19.40	16.77	17.03
Guiraud's Index of Lexical Richness	5.93	5.97	5.93	5.82
Accuracy				
% of Error-Free Clauses	48.74	47.17	50.89	46.33
% of Self- Repairs	7.54	2.20	1.45	5.61
Ratio of Repaired Errors to Unrepaired Errors	9.90	2.41	1.58	7.03
Fluency				
Unpruned Speech Rate A	126.23	122.11	122.73	125.01
Pruned Speech Rate B	115.99	104.97	112.65	117.87

Patterns of Syntactic Complexity and Percentage of Lexical Words for the four levels of tasks are shown diagrammatically in Figures 7 and 8 below.

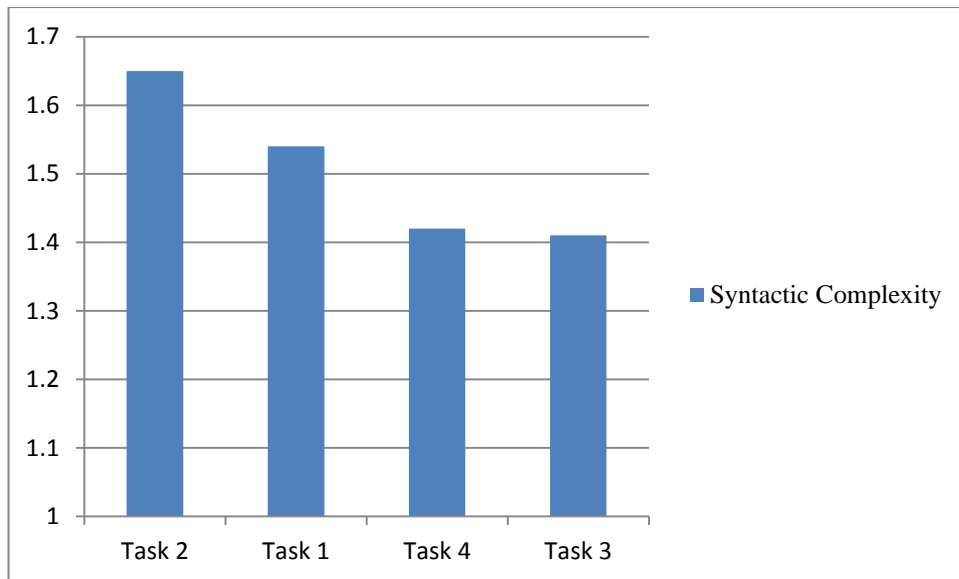


Figure 7: Means of Syntactic Complexity for the four levels of tasks.

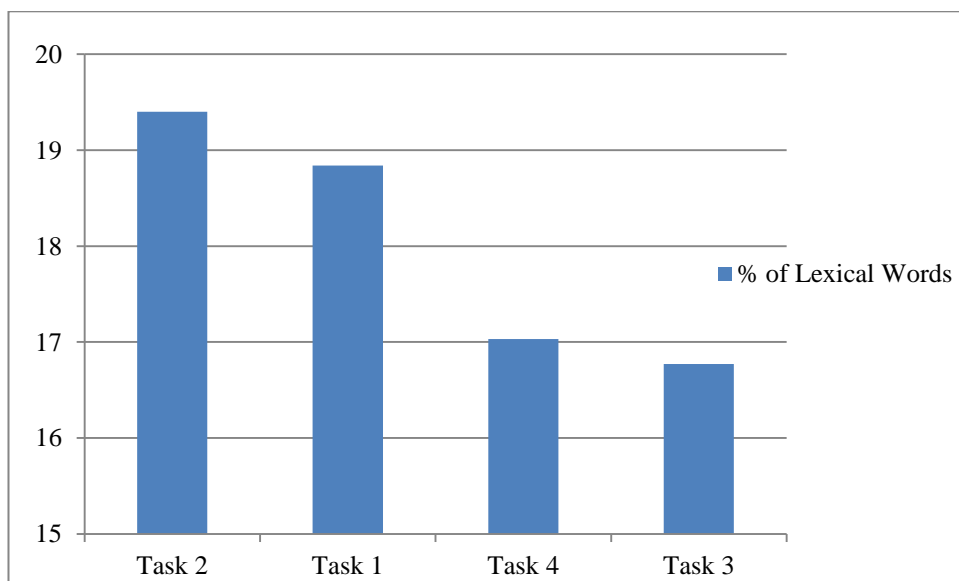


Figure 8: Means of Percentage of Lexical Words for the four levels of tasks.

In contrast, for Guiraud's Index of Lexical Richness, Tasks 1 and 3 had the same mean score, while Task 4 had the lowest mean score, however, this was not statistically different (see further discussion below) as shown in Figure 9 below.

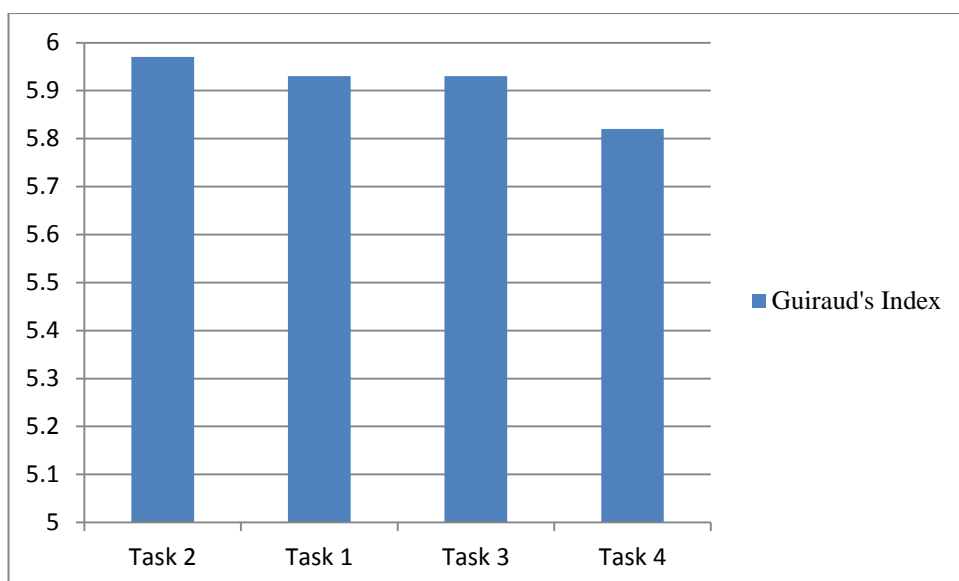


Figure 9: Means of Guiraud's Index generated by four levels of tasks.

Overall, these results suggest that the participants generated more complex syntactic constructions, lexically richer oral production, and more lexical words when they performed simple tasks (Tasks 1 and 2), especially when provided with planning time beforehand, than when they performed the two complex tasks (3 and 4). (This is examined in more detail in section 4.2).

With respect to **Accuracy** measures, Task 3 (the planned complex task) had the highest mean of Error-Free AS-Units among the four tasks, while the second highest mean was for the planned simple task (Task 1), followed by both the unplanned tasks – simple and complex (Tasks 2 and 4). The pattern of Error-Free AS-Units in four versions of tasks is shown diagrammatically in Figure 10 below.

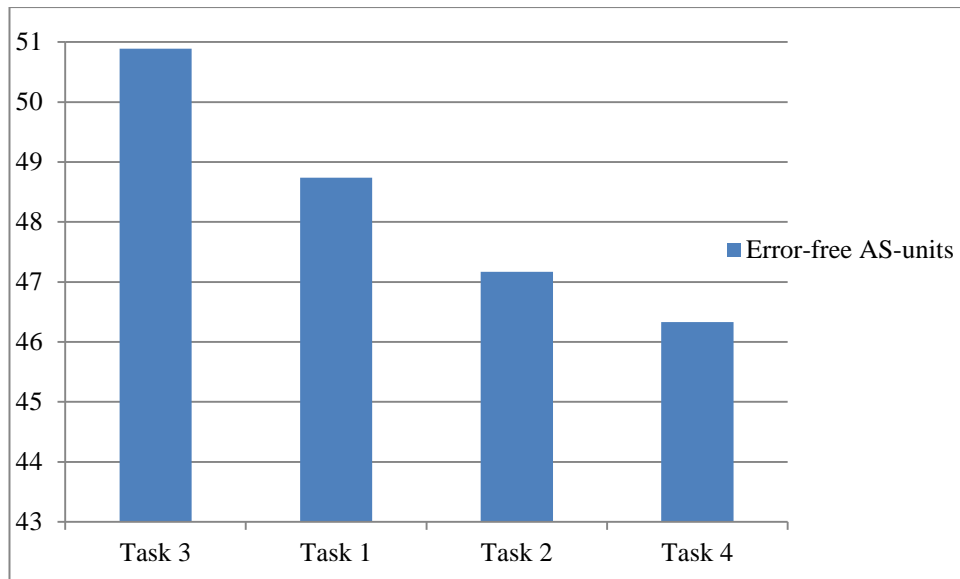


Figure 10: Means of Accuracy, Error-Free AS-Units, for the four levels of tasks.

However, Task 1 had the highest mean of Accuracy as measured by Percentage of Self-Repairs and Ratio of Self-Repaired to Unrepaired Errors followed by Tasks 4, 2, and 3, respectively as shown in Figure 11 below. Together these results suggest that planning (regardless of whether it is for a simple or complex task) led the participants to generate slightly more accurate oral production, but, that planned simple and unplanned complex tasks (Tasks 1 and 4) triggered the participants to self-repair and to do so more effectively.

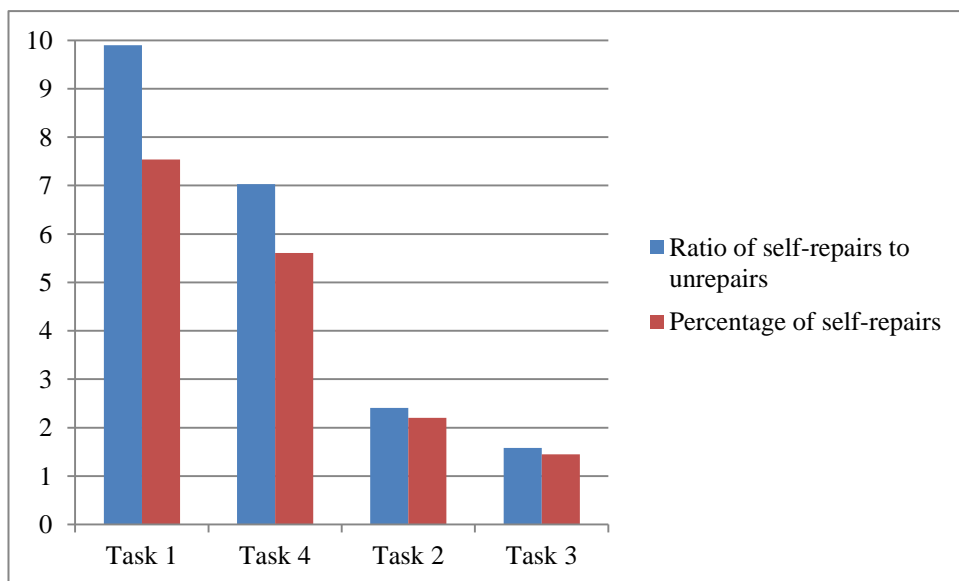


Figure 11: Means of Accuracy, Ratio of Self-Repairs to Unrepaired Errors and Percentage of Self-Repairs for the four levels of tasks.

For **Fluency** measures, Task 1 generated the highest mean as assessed by Speech Rate A. This was followed by Tasks 4, 3, and 2 respectively. However, with respect to the Fluency of Speech Rate B, Task 4 produced the highest mean, followed by Tasks 1, 3, and 2. Once again the pattern that emerges is one where planned simple and unplanned complex resulted in greater fluency than the two other task configurations. This is shown diagrammatically in Figure 12 below.

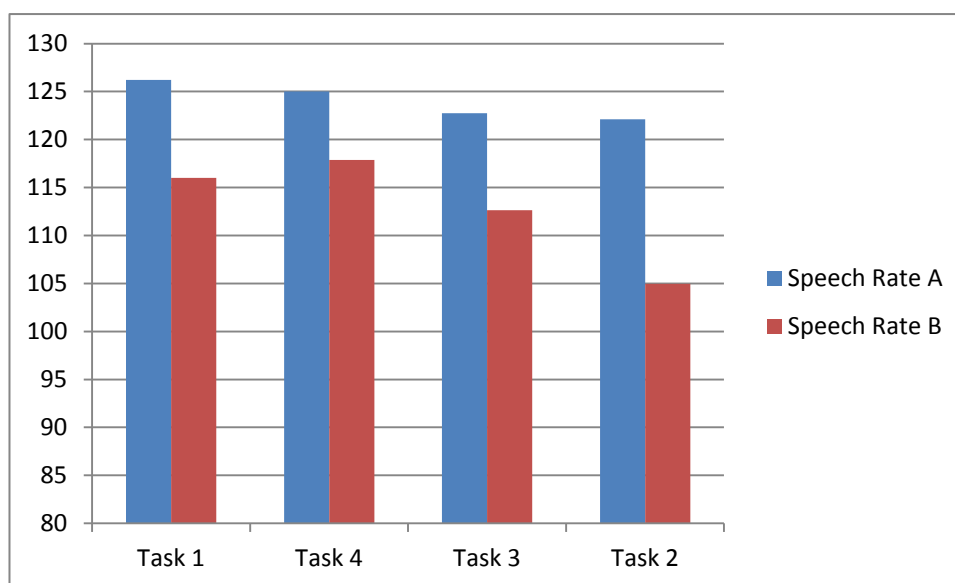


Figure 12: Means of Fluency Speech Rate A and Speech Rate B for the four levels of tasks.

The effect of the four levels of task difficulty on CAF measures was tested using Repeated Measures of Analysis of Variance (ANOVA). Multivariate analysis was then used to examine the main effects size of the measures for the four levels of the tasks. This was done using Wilks' Lambda because the value of sphericity in the current study was violated (Pallant, 2007, p. 255). These results are reported in Table 7 below.

Table 7: Repeated measures ANOVA: main effects of different levels of tasks for CAF measures

Measures		Wilks' Lamda	F-value	p-value	η
Complexity	Syntactic complexity	.507	15.894	.000*	.493
	Percentage of Lexical words	.523	14.896	.000*	.477
	Guiraud's Index	.934	1.162	.334	.066
Accuracy	Error-Free-AS-Unit	.825	3.465	.023*	.175
	Percentage of Self-Repairs	.660	8.398	.000*	.340
	Ratio of Repaired to Unrepaired	.693	7.237	.000*	.307
Fluency	Speech Rate A	.966	.566	.640	.034
	Speech Rate B	.723	6.244	.001*	.277

* $p < 0.05$, η = Partial Eta Square

As can be seen there was a statistically significant effect for six out of eight aspects of CAF measures ($p < 0.05$), namely, two of the Complexity measures (Syntactic Complexity and Percentage of Lexical Words), three Accuracy measures (Error-Free AS-Units, Percentage of Self-Repairs, and Ratio of Self-Repaired to Unrepaired Errors), and one Fluency measure (Speech Rate B). In contrast, there were no statistically significant differences between the tasks for the two CAF measures, Guiraud's Index of Lexical Richness and for Fluency as measured by Unpruned Speech Rate A. For those measures that were significantly different for the four tasks, the results show a large effect size (i.e., the values of Partial Eta Squared obtained from the multivariate tests were higher than .14). From this, it does appear that planning time and the number of elements affected the learners' performance, but perhaps not in expected ways.

However, it must be noted that these findings did not show which tasks contributed to the statistically significant difference on six out of eight CAF measures ($p < 0.05$). Therefore, the comparisons of each pair for the four levels of task complexity are

considered in more detail in the following sections in order to examine the contribution of each version of the tasks on CAF.

These findings do suggest that the four levels of dialogic task complexity manipulated simultaneously along with planning time and the number of elements only partly support the Cognition Hypothesis. According to the Cognition Hypothesis, Task 3 (the planned complex task) would normally be predicted to result in more fluent oral production compared to Task 4 (the unplanned complex task). Similarly, based on the Cognition Hypothesis it was predicted that Task 1 (the planned simple task) would result in higher Guiraud's Index ("more varied lexis") compared to Task 2 (the unplanned complex task). However, the evidence from the measure of Speech Rate A indicates that actually Task 3 generated less fluent oral production, and when measured by Guiraud's Index Task 2 resulted in higher fluency scores compared to Task 1. These results and the inconsistency with the Cognition Hypothesis are considered in more detail in the following sections.

4.3 Comparing Planned and Unplanned Tasks

This section reports on the apparent effect (as measured by CAF) of the planned and unplanned conditions for both simple tasks (Tasks 1 and 2) and complex tasks (Tasks 3 and 4). First the results of Complexity measures for the comparison of Tasks 1 and 2, and Tasks 3 and 4 are presented in Table 8 below.

Table 8: Mean differences of planned and unplanned tasks for Complexity Measures

Comparison	Syntactic Complexity	Percentage of Lexical Words	Guiraud's Index
Planned Simple Task (Task 1) vs. Unplanned Simple Task (Task 2)	-0.11*	-0.56	-0.40
Planned Complex Task (Task 3) vs. Unplanned Complex Task (Task 4)	- 0.01	-0.26	0.11

*The mean difference is significant at the 0.05 level.

As described in Table 8 above (p. 75) in terms of Complexity, regardless of whether the task was simple or not, the participants produced more complex syntactic constructions when they performed the unplanned tasks (Tasks 2 and 4) rather than when they performed the planned tasks (Tasks 1 and 3). That is, by increasing difficulty, and by not providing planning time (– planning time), this current cohort generated oral production with more complex syntactic constructions (Syntactic Complexity). However, as shown in Table 3 above, Syntactic Complexity was only statistically significantly different for Task 1 compared to Task 2 (0.11, $p < 0.05$) not for Task 3 compared to Task 4 (0.01, *n.s.*).

Similar results occurred for Lexical Complexity. Specifically, the participants seemingly generated slightly more complex oral production as shown by the higher percentage of lexical words for the unplanned simple task (Task 2) than for the planned simple task (Task 1) (– 0.56, *n.s.*). They also produced a slightly higher percentage of lexical words for the unplanned complex task (Task 4) than for the planned complex task (Task 3) (– 0.26, *n.s.*). However, it must be noted that these differences were not statistically significant.

Although Guiraud's Index of Lexical Richness showed neither a significant difference between the tasks, nor a large effect size, the pattern of production is interesting as a point of contrast to the other Complexity measure results. The participants generated lexically richer oral production, as indicated by their higher Guiraud's Index score, when they performed the unplanned simple task (Task 2) compared to the planned simple task (Task 1). On the other hand, their oral production was lexically richer for the planned complex task (Task 3) than for the unplanned complex task (Task 4). Despite the lack of significance, this does show that, for complexity at least, there is a dynamic, albeit subtle relationship between task difficulty and planning time.

In general, these findings indicate that providing ten minute planning time for both the simple and complex (dialogic) tasks did not appear to help the participants to generate more complex oral production. In fact, generally there appears to be a 'negative' effect from providing planning time on the three complexity measures viz. Syntactic Complexity, Percentage of Lexical Words, and Guiraud's Index of Lexical

Richness. One exception is Task 3 (the planned complex task) which did generate lexically richer oral production (higher Guiraud's Index) compared to Task 4 (the unplanned complex task) though the difference is not statistically significant. Further, this finding is consistent with that of Gilabert (2005) who found that providing ten minute planning time when the task is more complex within the resource-directing dimensions (i.e., many elements to compare) enables learners to generate lexically richer oral production.

Although the current findings largely contradict the Cognition Hypothesis which predicts that increasing complexity along the resource-dispersing dimension (i.e., no planning time), will result in a negative effect on the aspects of linguistic complexity (Robinson, 2003, 2005), the findings are consistent with studies by Gilabert (2005); Mehnert, (1998); and Yuan and Ellis (2003, 2005). As in Gilabert's study, the current findings suggest that providing ten minute planning time results in more complexity in terms of oral production within the resource-directing dimensions (i.e., many elements to compare). Like Mehnert, (1998) and Yuan and Ellis (2003, 2005) the current findings also found no significant difference between planned and unplanned conditions, at least for some measures of syntactic complexity.

At the same time the current findings do contrast to those of the previous studies which indicated the positive effect of planning time on learners' language performance – at least in terms of complexity measures (Ortega, 1999; Philp, Oliver, and Mackay, 2006). Clearly there is still a great deal more research to do in different contexts.

There are a number of possible reasons underlying the 'negative' effect of manipulating planning time on the three complexity measures in this study. As predicted by the Cognition Hypothesis, "the manipulation of the resource-dispersing dimension - with or without planning time - does not direct learners to any particular aspects of language code which can be used to meet the additional task demands" (Robinson, 2005, p. 7). Another possibility is that the participants might not have planned maximally what they intended to say, therefore, they could not "access their established repertoires of L2 knowledge" (Robinson, 2005, p.7). Based on these reasons, it is possible that planning time could facilitate learners to generate more

complex oral production, but only when they have adequate target language knowledge related to the task they are going to perform. Conversely, planning time may not facilitate learners to generate more complex oral production if their L2 knowledge is not adequate. As Ellis (2005) argues, “it is essentially a problem solving activity; it involves deciding what linguistic devices need to be selected in order to affect the audience in the desired way” (p. 3).

It is also possible that the repetition effect of performing the previous tasks (i.e., familiarity with task) leads to more oral complex production from participants with the unplanned simple and complex tasks. This is supported by several participants’ views about the unplanned tasks for both the simple and complex tasks which they regarded as “easy” due to familiarity with the previous task (e.g., “I think easy because I have done the similar task, so I feel usual to make conversation”) (This is discussed in more detail in Chapter Five). The current findings also provide additional support for Bygate’s (1999) argument that task repetition leads to better language performance in terms of CAF because it provides learners with “the time and awareness to shift attention from message content to the selection and monitoring of appropriate language” (p.41). The shift in attention might enable the participants to retrieve more of their current L2 knowledge. This circumstance might ultimately contribute to more complex syntactic constructions, a higher percentage of lexical words, and lexically richer oral production. Therefore, it is reasonable to assume that repetition or familiarity with performing tasks does impact on the contribution of planning time.

In short, the ‘negative’ effect of manipulating planning time on the Complexity of learners’ performance may be due to i) the nature of the cognitive factors of the resource-dispersing dimensions (i.e., planning time) as predicted by the Cognition Hypothesis, ii) learners’ factors (i.e., proficiency), or iii) the effect of another cognitive factor of the resource-directing dimensions (i.e., familiarity or prior knowledge).

With respect to Accuracy, the results of different means for comparison of Tasks 1 and 2, and Tasks 3 and 4 are presented in Table 9 below.

Table 9: Mean differences of planned and unplanned tasks for three Accuracy Measures

Comparison	Error-Free AS-Units	Percentage of Self-Repairs to Number of Errors	Ratio of Self-Repaired to Unrepaired
Planned Simple Task (Task 1) and Unplanned Simple Task (Task 2)	1.58	5.34*	7.49*
Planned Complex Task (Task 3) and Unplanned Complex Task (Task 4)	4.55*	- 4.16*	-5.45*

*The mean difference is significant at the 0.05 level

As can be seen from Table 9 above, the planned condition in both the simple and complex tasks (Tasks 1 and 3) generated more accurate oral production as measured by Error-Free AS-Units compared to the unplanned simple and complex tasks (Tasks 2 and 4). However, the difference for Accuracy was only statistically significant between Tasks 3 and 4 (4.55, $p < 0.05$), not between Task 1 and 2 (1.58, *n.s.*). When Accuracy is measured by Percentage of Self-Repairs, the results indicate that the planned simple task (Task 1) generated more accurate oral production, as shown by a higher percentage of self-repairs than the unplanned simple task (Task 2) (5.34, $p < 0.05$). On the other hand, the participants produced less accurate oral output as indicated by a lower percentage of self-repairs, when they performed the planned complex task (Task 3) than when they performed the unplanned complex task (Task 4) (- 4.16, $p < 0.05$). Further, these differences were statistically significant. This is a similar pattern to that shown in Figure 11, page 75.

For the Ratio of Repaired to Unrepaired Errors the pattern of Accuracy was also similar. Specifically, the planned simple task (Task 1) appeared to generate more accurate oral production than the unplanned simple task (Task 2) (7.49, $p < 0.05$). On the other hand, the planned complex task (Task 3) resulted in less accurate oral output than the unplanned complex task (Task 4) (-5.45, $p < 0.05$). Again, the

differences in the Ratio of Repaired to Unrepaired Errors were statistically significant. Overall, therefore, it appears that the planned simple task generated the most accurate oral production, and depending on the measures used, planning time could increase or decrease Accuracy in the complex tasks. Once more these show an interesting and dynamic relationship between accuracy of production, planning time and task difficulty.

These findings suggest that providing learners with ten minute planning time generally leads them to produce more accurate oral production when the task is simple, that is, within the resource-directing dimensions (+ *few elements*). As predicted by the Cognition Hypothesis, this is because, although providing planning time does not “facilitate new form-function mappings in the L2”, it enables learners to access to their established L2 knowledge. Therefore, lack of planning time prior to performing tasks may “create problems for learners attempting to access their current repertoire of L2 knowledge” (Robinson, 2005, p.7). The findings of this research are also largely consistent with the previous studies with monologic task conditions (Foster and Skehan, 1996; Mehnert, 1998; Ortega, 1999; Yuan and Ellis, 2003, 2005; Ahmadian and Tavakoli., 2010), that is, providing planning time results in more accurate oral production. These findings are also in agreement with Ortega’s (1999) argument that planning time enables learners to focus their attention on form, engage cognitive and metacognitive strategies, and monitor their language production.

However, the findings of this current study also show that providing ten minute planning time does not result consistently in increased Accuracy outcomes. For example it led to a decrease in Accuracy as shown by a lower Percentage of Self-Repairs and Ratio of Self-Repaired to Unrepaired Errors, especially when the task was more complex (more cognitive engagement). Again, with respect to these two measures it appears that planning time might not significantly help learners in attempting to generate more accurate oral production if their current repertoires of L2 knowledge are not adequate to repair the errors they make. It is also possible that providing learners with planning time disperses the participants’ attention, and prevents them from making repairs when the task is more complex.

In terms of Fluency, the results of the comparisons of Tasks 1 and 2, and Tasks 3 and 4 are shown in Table 10 below.

Table 10: Mean differences of planned and unplanned tasks for two Fluency Measures

Comparison	Unpruned Speech Rate A	Pruned Speech Rate B
Planned Simple Task (Task 1) and Unplanned Simple Task (Task 2)	4.11	11.02*
Planned Complex Task (Task 3) and Unplanned Complex Task (Task 4)	-2.28	-5.22

*The mean difference is significant at the 0.05 level

As can be seen from Table 10 above, when measured using Speech Rate B, the participants were more fluent in the planned task (Task 1) than in the unplanned simple task (Task 2), and the difference was statistically significant (11.02, $p < 0.05$). However, there was no significant difference between Tasks 3 and 4 for Speech Rate B, (-5.22, *n.s.*). It would seem that providing planning time, at least for simple tasks may increase Fluency, but when the task is complex, planning time may have the opposite or no effect at all.

Thus, the findings of this study only partly confirm the Cognition Hypothesis, which predicts that increasing complexity within planning time decreases fluency. These findings are also only in partial agreement with those found in the previous studies investigating the role of planning time on CAF (Foster and Skehan, 1996; Mehnert, 1998; Ortega, 1999; Yuan & Ellis, 2003, 2005; Gilabert, 2005; Ahmadian & Tavakoli, 2010).

Furthermore, the decrease of fluency with the planned complex task (Task 3) suggests that providing planning time may not lead learners to generate more fluent oral production when the task is a complex one within the resource-directing dimension, such as when there are many elements to discuss. This finding is in line

with participants' perceptions in which the planned complex task (Task 3) was most frequently regarded as the most difficult task due to it being complex (i.e., there were many elements to compare). (Again this is discussed in detail in Chapter Five). Interestingly, the unplanned complex task (Task 4), which is regarded as the most complex task as predicted by the Cognition Hypothesis, generated more fluent oral production than Task 3. Again, it is possible that the increase in fluency for Task 4 was due to the impact of task repetition, as was also found in the study of Bygate (1999) or to the "familiarity with the tasks" as argued by Skehan (1998). According to Bygate, learners no longer pay more attention to the "message content" of the tasks but rather to "the selection and monitoring of appropriate language" (p.41) when they are familiar with the model of conversation. Skehan (1998) argues that learners will perform a task with which they are familiar more easily. These findings are also in line with the participants' views about their familiarity of performing the previous task as mentioned frequently for the unplanned complex task (Task 4) (e.g., "It's easy because I have done the tasks before and practice with tasks before makes me easy to do the task" – see Familiarity Chapter Five).

4.4 Comparing Simple and Complex Tasks

This section presents the comparison of CAF measures for the simple and complex tasks (+/– *few elements*), namely, Tasks 1 and 3, and Tasks 2 and 4. These are presented in Table 11 below.

Table 11: Mean differences of simple and complex tasks for Complexity measures

Comparison	Syntactic Complexity	Percentage of Lexical Words	Giuraud's Index
Planned Simple Task (Task 1) vs Planned Complex Task (Task 3)	0.13*	2.10*	.003
Unplanned Simple Task (Task 2) vs Unplanned Complex Task (Task 4)	0.23*	2.37*	0.15

*The mean difference is significant at the 0.05 level

As can be seen from Table 11 above, the syntactic constructions as measured by Syntactic Complexity for the simple and complex tasks in both the planned and unplanned conditions were statistically significantly different. This is similar to the pattern shown in Figure 7, on page 77. These findings suggest that increasing difficulty with the number of elements for both the planned and unplanned conditions triggered the participants to produce less complex syntactic constructions.

As shown in the Table above, similar patterns also emerged when Complexity was measured by the Percentage of Lexical Words, as the planned simple task (Task 1) generated more complex oral production than the planned complex task (Task 3) (2.10, $p < 0.05$). The unplanned simple task (Task 2) resulted in more complex oral production than the unplanned complex task (Task 4) (2.37, $p < 0.05$). As indicated, the differences between them were statistically significant. These results show that the two simple tasks in both the planned and unplanned conditions enabled the participants to produce a higher percentage of Lexical Complexity. Thus it appears that increasing the difficulty according to the number of elements generated less complex oral production as evidenced by a lower percentage of lexical words.

With respect to Guiraud's Index of Lexical Richness, the two simple tasks (Tasks 1 and 2) in both the planned and unplanned conditions enabled the participants to generate lexically richer oral production than the complex tasks (Tasks 3 and 4). These findings seem to suggest that increasing difficulty with the number of elements (*– few elements*) resulted in lexically less rich oral output. However, as indicated, the differences were not statistically significant: Tasks 1 and 3 (0.003, *n.s.*); and Tasks 2 and 4 (0.150, *n.s.*).

To summarize, the participants appeared to generate more complex syntactic constructions and a higher percentage of lexical words when they performed the simple tasks than when they performed the complex tasks, for both the planned and unplanned conditions, suggesting that complex tasks, with an increased number of elements led to less complex oral production.

These findings suggest that increasing complexity with the number of elements for both the planned and unplanned conditions triggered the participants to generate less

complex syntactic constructions, a lower percentage of lexical words, and lexically less rich oral production. As predicted by the Cognition Hypothesis, the decrease of syntactic constructions is considered to be due to the nature of the interactive (dialogic) tasks, which prompts learners to produce “clarification checks” and “turn-taking” strategies, ultimately leading the participants to produce shorter sentences.

These findings are in agreement with the studies of Robinson (2001a) and Michel et al. (2007). That is, the participants produced less complex syntactic constructions when the task was a complex one (more cognitive engagement). However, the findings regarding oral production as measured by Percentage of Lexical Words and Guiraud’s Index contradict those found in the previous studies (Robinson, 2001a; Michel et al., 2007). The results of the current study revealed less complex oral production as indicated by a lower percentage of lexical words and lexically less oral production for the planned and unplanned complex (dialogic) tasks (Tasks 3 and 4) compared to the planned and unplanned simple (dialogic) tasks (Tasks 1 and 2). It is possible that differences in interactive factors, that is, dialogic tasks in the current study rather than monologic tasks as in Robinson’s (2001a) study may account for these different results. Robinson (2003, 2005) and Michel et al. (2007) argue that the dialogic (interactive) tasks, especially complex ones, are commonly characterized by highly interactional conversation (i.e., a lot of turn-taking and clarification requests). This condition may disperse the learners’ attention from what they have planned to say and, consequently, they produce simpler clauses as well as less varied lexis (Robinson, 2003, 2005; Michel et al., 2007).

In terms of Accuracy, the results of Accuracy measures for comparisons of Tasks 1 and 3, and Tasks 2 and 4 are presented in Table 12 below.

Table 12: Mean differences of simple and complex tasks for three Accuracy measures

Comparison	Error-Free AS-Units	Percentage of Self-Repairs to Number of Errors	Ratio of Self-Repaired to Unrepaired
Planned Simple Task (Task 1) and Planned Complex Task (Task 3)	-2.15	6.08*	8.32*
Unplanned Simple Task (Task 2) and Unplanned Complex Task (Task 4)	0.83	-3.41*	-4.62*

*The mean difference is significant at the 0.05 level

As can be seen from Table 12 above, although the planned complex (Task 3) and the unplanned simple task (Task 2) generated more accurate oral production in terms of Error-Free AS-Units compared to the planned simple (Task 1) and unplanned complex task (Task 4) respectively, the differences between the simple and complex tasks were not statistically significant.

With regard to Percentage of Self-Repaired to Errors, the planned simple task (Task 1) generated more accurate oral output than the planned complex task (Task 3) (6.08, $p < 0.05$). In contrast, the unplanned simple task (Task 2) generated less accurate oral production compared to the unplanned complex task (Task 4). Further, the differences were statistically significant (-3.42 , $p < 0.05$). These findings indicate that increasing difficulty with the number of elements (*–few elements*) within the planned condition (*+ planning time*) decreased Accuracy as shown by a lower percentage of self-repairs. Conversely, it increased Accuracy for the unplanned condition (*– planning time*). This is similar to the pattern shown for Figure 11, on page 79.

A similar pattern occurred for the results for Accuracy as measured by the Ratio of Self-Repairs to Errors: The planned simple task (Task 1) generated more accurate oral production than did the planned complex task (Task 3) (8.32, $p < 0.05$). On the other hand, the unplanned simple task (Task 2) resulted in a lower ratio of self-repairs than the unplanned complex task (Task 4) (-4.62 , $p < 0.05$). Again, the

differences were statistically significant. To summarize, there was no significant difference between the simple and complex tasks as measured by Error-Free AS-Units. On the other hand, as measured by the Percentage of Self-Repairs and the Ratio of Self-Repaired to Unrepaired Errors, the simple unplanned task resulted in greater Accuracy than the planned simple task, whereas with the complex tasks the planned condition produced significantly greater Accuracy than did the unplanned task.

These findings partly confirm the Cognition Hypothesis. That is, the complex task manipulated along the resource-directing dimensions (i.e., – few elements) led to an increase in the accuracy of language production. However, the increase in accuracy as measured by Error-Free AS-Units was only confirmed for Task 3 (as compared to Task 1). In contrast, Accuracy in terms of Percentage of Self-Repairs and Ratio of Repaired Errors only occurred for the complex task with the unplanned condition (Task 4) compared to the simple unplanned task (Task 2). These findings are, in the main, similar to those in the study by Michel et al. (2007), that is, complex dialogic tasks generated more accurate oral production as measured by Error-Free As-Units, and Ratio of Repaired Errors, but it produced less accurate oral output as evidenced by a lower Percentage of Self-Repairs.

With respect to Fluency, the comparisons of the simple and complex tasks within the planned and unplanned conditions (Tasks 1 and 3, and Tasks 2 and 4) are shown in Table 13 below.

Table 13: Mean differences of simple and complex tasks for two Fluency measures

Comparison	Unpruned Speech Rate A	Pruned Speech Rate B
Planned Simple Task (Task 1) and Planned Complex Task (Task 3)	3.50	3.34
Unplanned Simple Task (Task 2) and Unplanned Complex Task (Task 4)	2.90	–12.90*

*The mean difference is significant at the 0.05 level

Table 13 shows that the planned simple task (Task 1) appeared to trigger the participants to generate more fluent oral production as measured by Speech Rate A than the planned complex task (Task 3), but the result was not statistically significant (3.50, *n.s.*). In contrast, the simple task, when unplanned (Task 2) appeared to generate less fluent speech as measured by Speech Rate A than the unplanned complex task (Task 4), but again this was not statistically significant (−2.90, *n.s.*). These results indicate that increasing difficulty with the number of elements (− *few elements*) for the planned condition (+ *planning time*) does not affect the Fluency of Speech Rate A. The pattern is also repeated when Fluency is measured using Speech Rate B. Specifically, the planned simple task (Task 1) seemed to generate more fluent oral production than the planned complex task (Task 3), but again it was not statistically significant (3.34, *n.s.*). On the other hand, the unplanned simple task (Task 2) did result in significantly less fluent oral production as measured by Speech Rate B than the unplanned complex task (Task 4) (12.90, $p < 0.05$). Therefore, with respect to Fluency when no planning time is provided, it appears that complex tasks enable participants to generate more fluent oral production.

From the results described above it would seem that there is a complex interrelationship between the conditions of complexity (+/− number of elements) and planning. Further, these findings are largely in agreement with the study of Michel et al. (2007) that simple dialogic tasks, that is, with few elements to compare, have the potential to generate more fluent oral production.

Moreover, these findings partly confirm the Cognition Hypothesis, that is, increasing complexity along the resource-directing dimension by including many elements to discuss, generates less fluent oral production. Further, the statistically significant increase of fluency as measured by Speech Rate B for Task 4 (over Task 2) might be considered to be due to the familiarity of performing the previous tasks as Task 4 was performed in the last round. As previously stated, it is possible that the participants produce more fluent speech for the unplanned complex task (Task 4) than for the unplanned simple task (Task 2) because they have become familiar with performing a similar model of conversation. These findings are in line with Skehan's (1998) concepts of task difficulty, that is, learners' degree of familiarity with the

nature of tasks or the topic will contribute to their level of difficulty in performing tasks.

4.5 Comparing Planned Simple and Unplanned Complex Tasks and Unplanned Simple and Planned Complex Tasks

This section reports the results from comparisons of the planned simple and the unplanned complex tasks (Tasks 1 and 4), and the unplanned simple and the planned complex tasks (Tasks 2 and 3). In Robinson's (2003, 2005) terms as described in Chapter Three, these comparisons deal with Low Performative and Low Developmental Complexity (Task 1) compared to High Performative and High Developmental Complexity (Task 4), and High Performative and Low Developmental Complexity (Task 2) compared to Low Performative and High Developmental Complexity (Task 3).

The results of comparisons of mean between Tasks 1 and 4, and Task 2 and 3 for three Complexity measures are presented in Table 14 below.

Table 14: Mean differences between Tasks 1 and 4, and Tasks 2 and 3 for Complexity measures

Comparison	Syntactic Complexity	Percentage of Lexical Words (Lexical Density)	Giuraud's Index
Planned Simple Task (Task 1) vs. Unplanned Complex Task (Task 4)	0.12*	1.81*	0.11
Unplanned Simple Task (Task 2) vs. Planned Complex Task (Task 3)	0.24*	2.63*	0.04

*The mean difference is significant at the 0.05 level

When the Complexity measures were examined it was found that the planned simple task (Task 1) generated significantly more complex syntactic constructions (Syntactic Complexity) when compared to the unplanned complex task (Task 4)

(0.12, $p < 0.05$). The unplanned simple task (Task 2) also produced significantly more complex syntactic constructions than the planned complex task (Task 3) (0.24, $p < 0.05$). As shown, the differences are statistically significant. Similarly, the planned simple task (Task 1) generated more complex oral production as shown by a higher percentage of lexical words than the unplanned complex task (Task 4) (1.81, $p < 0.05$) and, likewise, the unplanned simple task (Task 2) led the participants to produce a higher percentage of lexical words than the unplanned complex task (Task 4) (2.63, $p < 0.05$). Once again the differences between Tasks 1 and 4, and Tasks 2 and 3 were statistically significant. Whilst the planned simple task (Task 1) appeared to trigger the participants to generate lexically richer oral production, as measured by Guiraud's Index of Lexical Richness, than the unplanned complex (Task 4) (0.11, *n.s.*) task –the result was not statistically significantly different. Similarly their oral production was not lexically richer in the unplanned simple task (Task 2) than the planned complex task (Task 3) (0.04, *n.s.*) because, as indicated, the differences were not statistically significant.

To summarize, the findings for Complexity measures suggest that the number of elements and in particular, Low Developmental Complexity (i.e., simple rather than complex tasks with fewer elements), may have a greater impact on the production of syntactic Complexity than does the Performative Complexity of the tasks.

Again, these findings may be due to nature of the interactive (dialogic) tasks as predicted by the Cognition Hypothesis, that is, complex (dialogic) tasks, (i.e., many elements to compare), trigger learners to produce more confirmation checks and clarification requests. This circumstance will disperse learners' attention while performing tasks which ultimately leads to less complex oral output.

Furthermore, it appears that manipulating task complexity within the resource-directing dimension (i.e., +/- few elements) affects complexity more than manipulating the aspects of the resource-dispersing dimensions (i.e., +/- planning time). This is evidenced by higher scores for Complexity measures in the comparisons of the simple and the complex task for both the planned and unplanned condition, that is, both the simple tasks within planned and unplanned conditions enabled the participants to produce more complex syntactic constructions, a higher

percentage of lexical words, and lexically richer oral output (although the differences were not statistically significant). This is in line with the prediction of the Cognition Hypothesis that the resource-directing dimension (cognitive/conceptual demands) requires attention, and working memory directs learners to focus on linguistic forms. Although tasks, which are made more complex along the resource-dispersing dimension (e.g., – planning time), require more engagement in attention and working memory, they do not direct learners to the features of linguistic forms required to perform the task (Robinson, 2001a, 2001b, 2003, 2005).

With respect to Accuracy, the results of comparisons of mean between Tasks 1 and 4, and Tasks 2 and 3 for three Complexity measures are presented in Table 15 below.

Table 15: Mean differences between Tasks 1 and 4, and Tasks 2 and 3 for Accuracy measures

Comparison	Error-Free AS-Units	Percentage of Self-Repairs to Number of Errors	Ratio of Self-Repaired to Unrepaired
Planned Simple Task (Task 1) and Unplanned Complex Task (Task 4)	2.41	1.92	2.87
Unplanned Simple Task (Task 2) and Planned Complex Task (Task 3)	– 3.72*	0.75	0.83

*The mean difference is significant at the 0.05 level

In terms of accuracy there was no difference between the planned simple task (Task 1) and the unplanned complex task (Task 4) (2.41, *n.s.*), as measured by Error-Free AS-Units . In contrast, however, the unplanned simple task (Task 2) did result in significantly less accurate oral production in terms of Error-Free AS-Units than the planned complex task (Task 3) (–3.72, $p < 0.05$). There was no difference with regard to the Percentage of Self-Repairs in the planned simple (Task 1) and the unplanned complex task (Task 4) (1.92, *n.s.*). Similarly, although the unplanned simple task (Task 2) appeared to generate a slightly higher percentage of self-repairs compared to the planned complex task (Task 3) (0.75, *n.s.*), the differences were not statistically significant. Again when Accuracy was measured by Ratio of Self-

Repaired to Unrepaired Errors, there was no difference between Task 1 and Task 4 (2.87, *n.s.*), nor between Task 2 and Task 3 (0.83, *n.s.*).

Therefore, for Accuracy, no differences according to the various conditions were found, with the only exception, as measured by Error-Free AS-Units, being High Performative and Low Developmental Complexity (Task 2) compared to Low Performative and High Developmental Complexity (Task 3) (i.e., complex, unplanned resulted in greater Accuracy than simple planned tasks).

This finding is in agreement with the Cognition Hypothesis, that is, “increasing complexity along the resource-directing dimensions, e.g., many elements to compare, can be expected to be stronger when the task is simultaneously simpler along one or more resource-dispersing dimensions e.g., availability of planning time” (Robinson, 2005, p. 7). It appears that increasing complexity (complex task) within the resource-directing dimension enabled the participants to generate more accurate oral production as measured by Error-Free AS-Units when the ten minute planning time was given prior to task performance. This means that the planned complex task (Task 3) appears to have had a much greater effect on the accuracy measures, particularly Error-Free AS-Units, but it had little or no effect on the other two Accuracy measures (i.e., Percentage of Self-Repairs and Ratio of Self-Repaired to Unrepaired Errors).

However, with respect to percentage of self-repairs and the ratio of self-repaired to unrepaired errors there was no difference between the planned and unplanned complex tasks (Tasks 3 and 4). Therefore these findings contradict the Cognition Hypothesis which predicts that complex tasks along the resource-directing dimension would generate more accurate language production.

In relation to Fluency, the comparisons between Tasks 1 and 4, and Tasks 2 and 3 are presented in Table 16.

Table 16: Mean differences between Tasks 1 and 4, and Task 2 and 3 for Fluency

Comparison	Unpruned Speech Rate A	Pruned Speech Rate B
Planned Simple Task (Task 1) vs. Unplanned Complex Task (Task 4)	1.21	−1.88
Unplanned Simple Task (Task 2) vs. Planned Complex Task (Task 3)	0.62	−7.68*

*The mean difference is significant at the 0.05 level

The results of this study indicate no difference in fluency between the planned simple task (Task 1) and the unplanned complex task (Task 4) as measured by Speech Rate A (1.21, *n.s.*). Similarly, there was no difference between the planned complex task (Task 3) and the unplanned simple task (Task 2) (0.62, *n.s.*).

Whilst there was no difference between the planned simple task (Task 1) and the unplanned complex task (Task 4) as measured by Speech Rate B (−1.88, *n.s.*), when the unplanned simple task (Task 2) was compared to the planned complex task (Task 3) there was a statistically significant difference (−7.68, $p < 0.05$). Therefore, for Fluency (and only for one measure) as with Accuracy, High Performative and Low Developmental Complexity (Task 2) compared to Low Performative and High Developmental Complexity (Task 3) were the only conditions to show a difference.

This finding is in line with the prediction of Robinson (2003, 2005) that there will be a greater effect for Low Performative and High Developmental Complexity (Task 3) than High Performative and Low Developmental Complexity (Task 2). That is, tasks designed to be more complex along the resource-directing dimension (i.e., *− few elements*) and simultaneously simpler along the resource-dispersing dimension (i.e., *+ planning time*) are predicted to have a stronger effect on CAF.

Furthermore, the different results of the two fluency measures (Speech Rates A and B) between the simple and complex tasks might be due to the following conditions. On the one hand, although the participants were first assigned two simple versions of

the dialogic task (Tasks 1 and 2), which are predicted by the Cognition Hypothesis to be easier than the complex tasks (Tasks 3 and 4), it was possible that participants were not familiar with the nature or the model of the tasks. Consequently, in such circumstances, as Bygate (1999) argues, participants pay more attention to the “message content” than the “selection and monitoring of appropriate language” (p.41). Therefore, this condition may result in the participants’ oral production including more repetition, self-corrections, and false-starts (as measured by Speech Rate A), even if the task is a simple one.

On the other hand, the participants were expected to be familiar with performing the two complex tasks because they had performed similar models of dialogue in the two previous simple tasks. In this circumstance, as Bygate claims, the participants may “shift their attention from message content to the selection and monitoring of appropriate language” (p.41). Accordingly, familiarity with the nature of the tasks, regardless of the degree of complexity, might have led them to generate more oral fluency as evidenced by a fewer number of repetitions, self-corrections, and false-starts (as measured by Pruned Speech Rate B). Simple tasks may have generated less fluent oral production compared to complex tasks when the participants were not familiar with the nature or model of tasks. Conversely, complex tasks may have enabled the learners to generate more fluent oral production when they became familiar with the models of the task. Therefore, the extent of familiarity with either the topic or the model of task appeared to mediate the effect of manipulating the resource-directing dimension (i.e., +/- *few elements*) and simultaneously simpler along the resource-dispersing dimension (i.e., +/- planning time).

4.6 Summary and conclusion: Comparisons of task conditions according to CAF measures

A summary of the CAF results based on learner performance of four tasks that were simultaneously manipulated within planning time (+/- planning) and the number of elements (+/- few elements) is presented in Table 17 below.

The findings of this study suggest that overall the manipulation of dialogic task difficulty (cognitive factors) does have an effect on the participants’ oral language

performance as measured by CAF. However, when the tasks are compared at the individual level, there were variable results. Together, the findings of this study only partly confirm the Cognition Hypothesis that the participants' oral production in terms of CAF cannot be simply predicted by the manipulation of planning time and the number of elements.

Table 17: Summary of the effect of planning time and the number of elements

Measures	Planned vs. unplanned	Simple vs. complex	Planned simple vs. unplanned complex	Unplanned simple vs. planned complex
Complexity				
<i>Syntactic:</i> AS-unit	Different (simple tasks only)	Different	Different	Different
<i>Lexical:</i> % of Lexical Words	No difference	Different	Different	Different
Guiraud's Index of Lexical Richness	No difference	No difference	No difference	No difference
Accuracy				
Error-Free AS- Units	Different (complex tasks only)	No difference	No difference	Different
% of Self-Repairs	Different	Different	No difference	No difference
Ratio of Repaired Errors to Unrepaired Errors	Different	Different	No difference	No difference
Fluency				
Speech Rate A	No difference	No difference	No difference	No difference
Speech Rate B	Different (simple tasks only)	Different (complex tasks only)	No difference	Different

Chapter Five

FINDINGS OF INDONESIAN PARTICIPANTS' PERCEPTIONS OF FOUR LEVELS OF TASKS

5.1 Overview

This chapter addresses Research Questions 2 and reports on the findings of the participants' views about the complexity of the four tasks. It includes participants' responses and how they perceive the four versions of the tasks that had been simultaneously manipulated according to planning time and the number of elements. A thematic analysis of the data led to these learners' perceptions being grouped into seven categories, each of which is described in detail below.

5.2 The Indonesian Participants' Perceptions of Task Complexity

The seven categories of perceptions to emerge from the data included: difficulty, stress, confidence, interest, motivation, learning opportunity, and dialogic nature of the tasks. As described in Chapter Three, the responses for each category included the participants' contradictory opinions for each category which for coding purposes were symbolized as (+) and (–) as shown in Appendix 10. The former (+) refers to learners' agreement about an issue regarding the task, while the latter (–) indicates their disagreement about the task. For example, when a learner said that the task was easy to perform, the response was coded + (i.e., agreement). On the other hand, when a learner perceived the task as difficult, her or his comment was coded – (i.e., disagreement). An overview of these results showing the number and percentage of participants from the total cohort agreeing or disagreeing about a particular issue is shown in Table 18 below.

Table 18: Participants' agreement and disagreement about the complexity of four tasks

No	Category	Task 1		Task 2		Task 3		Task 4	
		Percentage		Percentage		Percentage		Percentage	
		+	–	+	–	+	–	+	–
1	Difficulty	73	27	75	25	21	79	48	52
2	Relaxed/Stress	80	20	76	24	47	53	69	31
3	Confidence	67	33	78	22	58	42	86	14
4	Interest	98	2	86	14	88	12	83	17
5	Motivation	100	0	94	6	82	18	84	16
6	Learning Opportunities	100	0	100	0	96	4	96	4
7	Dialogic (Interactive-ness)	100	0	100	0	100	0	100	0

As can be seen from Table 18 above, in general, the planned and unplanned complex tasks (Tasks 3 and 4) were perceived as more difficult and stressful than the planned and unplanned simple tasks. In contrast, the participants' degree of confidence, interest, and motivation did not necessarily decrease when the tasks were more complex. Moreover, the participants reported that all four levels of task provided learning opportunities and that they enjoyed the dialogic nature (i.e., the interactive-ness) of the tasks, regardless of the levels of complexity. As such there is little evidence to indicate a strong relationship between the Indonesian participants' perceptions regarding confidence, interest, motivation, learning opportunities, dialogic nature of the tasks and the four levels of task difficulty. However, the findings do suggest that there is a close relationship between the participants' feelings of difficulty and their degree of stress, and, the difficulty of the tasks.

In this way the findings of the current study only partly agree with Robinson's (2001a, 2001b, 2003, 2005, 2007a) argument that task complexity (cognitive factors) should be the sole basis for sequencing pedagogical tasks. It does seem that the participants' perceptions of task difficulty also need to be taken into consideration. The comments and level of agreement amongst the participants for issues emerging in each of the categories are examined in more detail in the following sections of this chapter.

5.2.1 Task Difficulty

As can be seen in Table 18 above, the results indicate that both the planned and unplanned simple tasks (Tasks 1 and 2) were perceived as being easy by the majority of the Indonesian participants (73% and 75% respectively), as indicated by the following comments e.g.,

It's easy because we only compare two things and each of the things has the features of blackberry (Task 1).

It's easy because in the task there are pictures so it's easier to say from pictures (Task 1).

I think easy because it's similar to the previous task just describing and comparing a simple topic for us (Task 2).

It is easy I think because we just compare between Blackberry onyx white and Blackberry bold I think and we just give our argument why we choose the only of two (Task 1).

On the other hand, the two complex tasks (Tasks 3 and 4) were mostly regarded as difficult (79% and 52% respectively), as evidenced in the following comments e.g.,

It is difficult because we have six different types compared with the previous task with only two types (Task 3).

I think it's difficult because there are a lot of pictures and information so it's difficult to make sentences which have relationship (Task 4).

These results provide some supporting evidence about the difficulty of the four levels of the tasks. Interestingly, the unplanned simple task (Task 2) was more frequently mentioned as an easy task than the planned simple task (Task 1). Similarly, the unplanned complex task (Task 4) was more frequently regarded as easy compared to the planned complex task (Task 3). Thus it appears that giving planning time does not necessarily lead the participants to perceive the tasks as being easier.

When the participants were asked what aspects of the task they found easy or difficult to perform, they gave varied responses as shown in Table 19 below.

Table 19: Participants' perceptions about the of difficulty of four levels of tasks

No	Perceptions	Percentage			
		Task 1	Task 2	Task 3	Task 4
	Reasons for feeling tasks were easy				
1	Planning time	2	-	-	-
3	Familiarity with the topic	24	28	2	14
4	Interest in the task	6	10	-	-
5	Amount of Information provided into task	14	8	-	-
6	Provided with pictures	12	4	-	-
7	Familiarity with the previous tasks	-	18	-	20
8	Having dialogue with a friend	2	-	-	2
9	Lack of numerical numbers	-	2	-	-
	Reasons for feeling tasks were difficult				
1	Problems with language	18	4	-	-
2	Complex tasks	-	-	82	46
3	No interest in the topic	-	7	-	2
4	Lack of familiarity with the topic	4	4	-	4
5	No planning time	-	8	-	-

A number of the participants found different aspects of the tasks helped simplify them. Further many of the comments from the participants suggest that they found it easier to perform the tasks due to factors other than task difficulty (a cognitive factor). For instance, some said that the amount of information provided about the tasks made the task easier (Task 1 – 14% and Task 2 – 8%), e.g.,

It is easy because we have to describe the sample provided with features we just add little vocabulary to describe the topic (Task 1).

There were others who found the tasks simple because they were interested in the topic (6% and 10% for Tasks1 and 2 respectively), e.g.,

The task is easy. I like the task because I like gadget (Task 1).

It is easy because the topic is more interesting than the first /blackberry/ can develop conversation because features are familiar and easy to understand (Task 2).

Pictures also enabled some of the participants to perform the tasks more easily, e.g.,

I think it is easy because we discuss in the task there are pictures so it is easier to what is it to say from the pictures (Task 1).

The pictures were commented on for Task 1 (12%) and to a smaller degree for Task 2 (4%), however, none of the participants mentioned these in relation to Tasks 3 and 4.

The participants also indicated that the interactive nature of the tasks, (i.e., being able to have a dialogue with a friend), made it easier for them to perform the tasks. This was especially so in relation to Tasks 1 and 4, e.g.,

It's easy because I have a partner (Task 1).

One issue on which there was considerable agreement was that familiarity with a topic made it was easy for the participants to perform the tasks.

It is quite easy to describe I think I have no problem to describe the feature because it is familiar to say we are familiar with these features you know everyone is familiar but you know everyone has different ability to say (Task 1).

I think it's easy for us to explain about or to make a daily conversation it's really easy I think because it's just like a daily conversation for us (Task 2).

They commented most frequently about familiarity in relation to the topics for Tasks 1 and 2 (24% and 28% respectively), but did so less often with regard to Tasks 3 and 4 (2% and 14%).

Several other participants said that they found it easy to perform both Task 2 (20%) and especially Task 4 (18%) because of their familiarity with doing the previous tasks (i.e., Tasks 1 and 3) as well as their familiarity with the topic, e.g.,

I think that the second task is easy maybe because I have done this task before; so in the second task I feel more usual to make conversation (Task 2).

Interestingly none of the participants commented on the “familiarity of doing the previous task” for Task 3. Based on their comments it does appear that repetition led the participants to perceive the tasks as being easier to perform, despite the tasks being complex (requiring greater cognitive engagement).

There were others who described how they valued having planning time (e.g., “It is not difficult because we have preparation before making conversation”). One participant commented specifically on planning time as an aspect that made performing the task easier:

I think it’s easy because the items are only two and the differences are not too big but overall it is not difficult because we have time to make preparation and just like what Putra said, the items are quite popular with us (Task 1).

Finally, some found the tasks easier because they did not involve anything related to number (Task 2 - 2%), (e.g., “It is easy because there are not many numerical numbers”).

Surprisingly a few participants found it easy to perform the more complex tasks when there were many elements to compare (14% for Task 3 and 12% for Task 4).

It’s easy because there are many specifications and there are six types of Blackberries (Task 3).

I think it’s easy because we have a lot of information. we have a lot of types of houses and because we are discussing with our friend and read it together so we can share our minds. That’s what makes it easy (Task 4).

This, however, was not common and in the interviews the planned complex task (Task 3) was frequently mentioned as the most difficult task (79%) followed by Task 4 (52%). Tasks 1 and 2 were nominated as difficult less frequently (27% and 25%). The participants described Tasks 3 and 4 as difficult because of the many elements they had to compare and they mentioned Task 3 more frequently (82%) than Task 4 (46%).

I think the task was difficult because we must choose many Blackberries and it made me confused (Task 3).

I think it's too difficult because there are so many pictures and much information in the pictures so I have difficulty to make sentences which have relationship between them (Task 4).

However, several participants regarded all four tasks as difficult. When the participants were asked why they found it difficult to perform the tasks, in addition to the complexity of the task, they mentioned such things as:

Problems associated with the language (Task 1 - 18%, and Task 2 - 4%)

It's difficult because my pronunciation I think is so bad I don't have many vocabularies I think, I lack my grammar /syntax/ (Task 1).

and lack of familiarity (4%) with the topic e.g.,

I feel frustrated because I don't have background about Blackberry and I have difficulty to compare six types of Blackberry and I feel confused (Task 3).

It's difficult because I have never rented a house before I don't have background knowledge (Task 2).

They had little or no interest in the topic (Tasks 1, 2, and 3, each 4%), found the tasks were too complex (82% - Task 3, and Task 4 - 46%), and that it was particularly difficult when no planning time (8%) was provided, e.g.,

It is a bit difficult, more difficult than the first because not given planning time to make preparation (Task 2).

I think the second task is a little bit more difficult than the first task because the lecturer did not provide us the time for making preparation and plan, so for the changing topic that we are going with our friend is a little bit hard to make (Task 2).

In short, the participants had problems in performing the four tasks not only because of the levels of task difficulty (cognitive factors), but also because of the other factors within the Triadic Componential Framework, (i.e., learner and interactive factors), and input i.e., pictures.

At the same time these findings do provide some evidence to suggest that, on occasions, the complexity of tasks (cognitive factors) does correspond to the degree of task difficulty (learner factors), so that the more complex tasks are, the more difficult learners will perceive them to be. In this way these findings are consistent with those of Robinson (2001a) and Gilabert (2005) who found that the participants' perceptions of task difficulty have a close relationship with the manipulation of task difficulty (a cognitive factor) and in this way does provide further support for the Cognition Hypothesis, that simple tasks along the resource-directing dimensions are expected to be easier for learners to perform as they are considered to involve less cognitive engagement compared to complex tasks. Conversely, complex tasks (increasing complexity), particularly within the resource-directing dimension (i.e., few elements) will be more difficult as they involve more cognitive engagement than simple tasks (Robinson, 2001a, 2001b, 2003, 2005).

Even so there is some contradictory evidence in the findings. Specifically, the unplanned simple task (Task 2) and the unplanned complex task (Task 4) were more frequently mentioned as easier tasks compared to the planned simple task (Task 1) and the planned complex task (Task 3). This is the opposite to what is predicted by the Cognition Hypothesis, specifically in the perception of the participants the manipulation of the resource-dispersing dimensions (i.e., planning time) does not impact on the degree of

task difficulty, particularly when compared to the manipulation of the resource-directing dimension (i.e., few elements).

One explanation for this may relate to the procedure followed in this study. It is possible that this trend is due to the participants being familiar with the procedure of the previous tasks as they frequently commented on this aspect of repetition for Tasks 2 and 4, while none mentioned the familiarity with the previous task for the planned simple and complex tasks (Tasks 1 and 3). That is, the familiarity of performing tasks may enable the participants to perform the tasks more easily, regardless of the degree of task complexity. This finding is in line with Skehan's (1998) concept of task difficulty (i.e., Cognitive Familiarity), especially "familiarity of task and discourse". According to Skehan, learners will be able to perform tasks with which they are familiar more easily than when they perform unfamiliar ones. The finding is also in agreement with Bygate's (1999) argument of "task repetition" viz, that learners will perform tasks more easily when they have repeatedly performed similar tasks. It appears that the more familiar the participants were with the previous model of tasks, the more easily they were able to perform the task whether it was a simple task or a complex one.

However, many of the participants also frequently perceived the two simple tasks as being easier due to them being familiar with the topics of the task, rather than the nature or model of the tasks. In this way their perceptions were in line with Skehan's (1998) concepts of "Cognitive Familiarity" - the "familiarity with the topic and its predictability, familiarity of discourse genre, and familiarity of task" (Skehan, 1998, p. 99) or Robinson's (2001a) "concept of task complexity" (cognitive factors), especially the aspect of the resource-depleting dimension (i.e., prior knowledge) where in both cases it is argued that familiarity with the content or the topic of tasks enables learners to perform tasks more easily (Skehan, 1998; Robinson, 2001a). The findings of this study suggest that the manipulation of task complexity (cognitive factors) within the resource-depleting dimensions (e.g., prior knowledge) may enable participants to perform tasks more easily. Certainly from the qualitative data (and the CAF results as described in Chapter Four) it does not appear that the increasing complexity of the number of elements and planning time led to any increase in the difficulty of tasks if the learners

were already familiar with them. Therefore it may be that two dimensions of task complexity, the resource-directing dimension (i.e., number of elements, reasoning demands, here and now, there and then) and the resource-dispersing (i.e., planning time, prior knowledge, single tasks), are interrelated rather than discrete and influenced by other factors such as familiarity.

Although the simple tasks were frequently regarded as easy, the participants did give a range of reasons for finding them so and in many cases this was due to aspects other than cognitive factors, such as interest in the tasks, having pictures, or having dialogue with a friend. Therefore, these findings suggest that, in the perception of the learners, the degree of task difficulty may be largely due to learner factors (i.e., learner affective and interactive factors, and, input such as pictures) rather than to cognitive factors, particularly when the task is simple.

Interestingly, while the participants indicated they had difficulty in performing the two complex tasks (Tasks 3 and 4) due mainly to them being more cognitively demanding, that is, more elements to compare, for some participants the two simple tasks (Tasks 1 and 2) were difficult because they felt they had problems associated with the language of the tasks. This aspect of language was not commented on for either the planned or unplanned complex tasks (Tasks 3 and 4). It might be that by this stage of the study they had developed sufficient language from their exposure to the tasks (Skehan, 1998; Robinson, 2001a), when undertaking the two simple tasks (i.e., Blackberry mobile phones and Houses for Lease). In this way the current findings and, in particular, the problems the participants identified that were associated with language, are consistent with the findings of Tavakoli (2009), namely that “linguistic demand” is one of the aspects underlying task difficulty. Hence, once again, it appears that the participants’ perception about the degree of task difficulty is not due simply to cognitive factors, but rather, is also due to “learner factors”.

Further, the difficulty the learners have with the language of the tasks maybe because they pay more attention to “the selection and monitoring of appropriate language”, rather than to the “message content” (Bygate, 1999, p.41). Although it is predicted that there is

less cognitive engagement with the simple tasks, they are not necessarily perceived as easy to perform by the participants because the language (e.g., vocabulary or syntax) required is beyond their current repertoire of L2 knowledge. Conversely, complex tasks might be regarded as easier if they have developed sufficiently from their experience with the previous tasks. In this way the participants might find it difficult to perform simple tasks if they lack familiarity with the topic, or they have problems associated with language.

Therefore, it does seem that simply increasing complexity along with either the resource-dispersing or the resource-directing dimensions does not have sufficient explanatory power with respect to the learners' perception of simple and difficult tasks. A number of other learner factors also appear to play a role.

5.2.2 Degree of Stress

As can be seen in Table 1 above, the majority of the participants said that they felt relaxed while performing the planned simple task (Task 1 - 80%), the unplanned simple task (Task 2 - 76%) and the unplanned complex task (Task 4 - 69%). However, they did feel stressed in the planned Complex Task (Task 3). Some also commented on their increasing stress when they had many elements to compare, and did so more frequently, as would be expected, with Tasks 3 and 4. However, "No planning time" was commented upon by only a few participants as contributing to their level of stress when performing the tasks, particularly for the unplanned simple task (Task 2). As such these findings do provide some evidence for the existence of a relationship between participants' degree of stress and the cognitive difficulty as manipulated according to the number of elements and planning time. However, in a similar way to the degree of difficulty, the participants gave a variety of reasons for why they felt more or less stressed by the tasks as outlined in Table 20 below.

Table 20: Participants' perceptions about the stress tasks

No	Perceptions	Percentage			
		Task 1	Task 2	Task 3	Task 4
	Reasons for being confident in performing tasks				
1	Planning time	2	-	2	-
2	Simplicity	2	4		
3	Complexity	-	-	4	12
	Familiarity with the topic	32	24	16	24
5	Interest in the task	12	10	6	4
6	Amount of information provided into task	-	6	-	-
7	Pictures	4	6	-	-
8	Familiarity with the previous tasks	-	12	8	22
9	Conversation with a friend	16	12	6	4
10	Source of motivation for learning English	6	2	2	-
11	Lack of numerical numbers	-	2	-	-
12	Absence of teachers	8	2	2	-
	Reasons for NOT being confident in performing tasks				
1	Problems with language	20	10	4	6
2	Complexity	-	-	46	24
3	Lack of familiarity with the topic	-	4	4	2
4	Lack of planning time	-	8	-	-

As indicated above, and in contrast to what would be expected based on Robinson's Cognition Hypothesis, only a few participants (Task 1 - 2% and Task 3 - 2%) said that planning time decreased stress levels:

I felt relaxed because it's easier for me to produce words because I have a chance before making conversation (Task 3).

In fact, the majority of the participants regarded the planned complex task (Task 3) as the most stressful of the tasks.

However, a few participants did suggest that they felt relaxed when doing Tasks 1 (2%) and 2 (4%) due to their simplicity, e.g.,

At first I feel nervous too because when we are told we don't know what to do but now I feel relaxed because the task is just comparing between two mobile phones (Task1).

As a corollary to this, for the two versions of the complex tasks (Tasks 3 and 4) some of the participants described these as “stressful” due to the many elements requiring comparison (i.e., being more complex) (4% and 12%), e.g.,

I'm frustrated maybe because this task has six types and I'm confused to choose this one and I don't know about Blackberry (Task 3).

Hence, these findings suggest that the participants' perceptions do align in some way to the levels of task difficulty. However, this was not consistently the case and it does appear that a number of factors beyond the complexity of the tasks (based on planning time and number of elements) contributed to the participants feeling either relaxed or stressed, e.g.,

I'm relaxed with pictures because I can imagine about them and I will speak up more (Task 2).

Again, several participants also agreed that being familiar with the tasks, by performing them previously, meant they felt less stressed and more relaxed. This occurred even when a task was more complex, but when the topic was one with which they are familiar.

I feel really relaxed because in this task we know Blackberries the topic is familiar in our life today (Task 1).

I feel relaxed because yeah the same as the previous reason because I have done the previous task so it will be easier (Task 2).

Familiarity with the topic and familiarity with the previous tasks is what Robinson (2001a, b) refers to as more or less prior knowledge, and it is one of the aspects within what he describes as the resource-dispersing dimension. In this study, based on the

qualitative evidence, it did seem that prior knowledge plays a more dominant role in decreasing participants' degree of stress rather than the manipulation of both the number of elements (the resource-directing dimension) and planning time (the resource-dispersing dimension). That is, even though the tasks were sequenced according to cognitive engagement, stress seemed to be more related to the issue of familiarity (or lack thereof). In the current study, the stress caused by a lack of familiarity with the topic was described in relation to Tasks 2, 3, and 4, but none of them commented on this aspect for Task 1. Although the topics of Tasks 1 and 3 are similar (Blackberries hand-phone brands), the models of Blackberry hand-phones of Task 1 are different from those of Task 3 and this may have contributed to the level of participant stress.

Difficulty of performing tasks due to lack of familiarity with the topic is consistent with the findings by Nunan and Keobke (1995), that learners will find it difficult to perform the task when they lack familiarity with the topic. This suggests that familiarity with the topic of the task helps learners perform the task more easily although it is cognitively more demanding.

Some of the participants also said that they felt stressed when performing the tasks because of problems associated with the language and the complexity of the task (many elements to compare), and planning time. Interestingly, the problems associated with language occurred more frequently with the two simple tasks than the two complex ones. These problems might simply be because the two simple tasks were performed before the complex ones; therefore, once they had done so the participants were more familiar with the language required for the complex tasks. As previously noted, it is possible that in this circumstance, learners, as Bygate (1999) argues, no longer pay more attention to the "message content" of the tasks but rather to "the selection and monitoring of appropriate language" (p. 41) when performing the two complex tasks. Therefore, these results suggest that the participants' degree of stress is not only related to increasing task complexity (cognitive factors), especially as manipulated by increasing the number of elements and planning time, but also due to task difficulty associated with learner factors (i.e., problems associated with language). However, as with task difficulty the impact of these factors does seem to have a dynamic relationship.

In terms of the participants' comments, other conditions, such as interactive factors (e.g., having dialogue with a friend) and learner factors (e.g., their interest in the topic) were more frequently regarded as being able to make participants feel relaxed than the manipulation of task complexity (cognitive factors) either within the resource-directing (i.e., simple task or comparing two) or the resource-dispersing dimension (i.e., planning time). For instance, some participants noted, with respect to interactive factors, that working together, especially without the presence of a teacher was not only motivating, but reduced their level of stress:

I think I feel relaxed because we are comparing and discussing with our friends so it's not frustrating (Task 2).

I think I'm relaxed because we were not attended by the teacher (Task 3).

I feel relaxed because it can help me to speak more. I felt more relaxed in this class than when I study in that class /regular class/ because in this class we're free to speak about everything. I'm not thinking about grammar. I'm not worried that the lecturer would be angry with me because of the problems of grammar /syntax/ (Task 1).

It's very enjoyable. I think same with her. We are not afraid to say /whether/ it's wrong. If it's false but we just say it's my words and I don't think I'm afraid if my grammar is so bad. It's not very good for people to hear it and I think I will say what I want. I'm free to say it without grammar (Task 1).

In summary, there was some alignment with the participants' degree of stress and the level of task difficulty, but a number of other factors also appeared to contribute to the participants feeling either stressed or relaxed about performing the tasks. The findings of this study suggest that the participants' degree of stress in performing tasks cannot be simply predicted by the manipulation of task complexity (cognitive factors) because it also includes other aspects of Robinson's (2001a, b, 2003, 2005) Triadic Componential Framework, viz., "interactive factors" (having dialogue with a friend), and "learner factors" (interest in the topic, problems associated with language, etc.,).

Although Robinson argues that learner factors cannot be predicted in advanced, this does not necessarily mean that other aspects in sequencing tasks should be neglected as individual differences inevitably exist within learner factors. Therefore, learners need to be provided with tasks or materials that are designed on the basis of not only cognitive factors, but also based on other factors, such as “interactive factors” (e.g., dialogic tasks), learner factors (e.g., interest in the topic) and other factors (e.g., provided with pictures) in accordance with their individual differences and needs. By being provided with these types of tasks, learners are more likely to be encouraged to learn English as these tasks also accommodate their needs. This circumstance may ultimately lead to their success in learning English. As Dornyei (2005) argues, individual differences provide an important indicator in contributing to the success of second language learning.

5.2.3 Degree of Confidence

The majority of the participants said that they felt confident and successful in performing the four tasks. They indicated that they thought they did the tasks well. This appeared to be particularly the case for the unplanned simple task (Task 2 - 78%) and the complex task (Task 4 - 86%), but slightly less so for the planned simple task (Task 1 - 67%) and planned complex task (Task 3 - 58%). Although, several participants did indicate that their confidence was affected by factors related to task complexity, specifically when performing complex tasks that had many elements to compare (Tasks 3 - 32% and Task 4 - 10%). A lack of planning time was also mentioned, but only for Task 2 and not with great frequency (4%). Overall, there is little alignment between these results and the four levels of task difficulty. The reasons underlying both learners’ confidence, or lack of confidence, in performing four versions of tasks are presented in Table 21 below.

Table 21: Students' perceptions about their confidence with tasks

No	Perceptions	Percentage			
	Reasons for confidence in performing task	Task 1	Task 2	Task 3	Task 4
1	Simple task	2	4	-	-
2	Complex task	-	-	10	12
3	Familiarity with the topic	24	22	10	26
4	Interest in the task	-	8	2	-
5	Amount of information provided into task	6	4	-	4
6	Pictures	2	4	-	-
7	Familiarity with the previous tasks	-	-	4	-
8	Conversation with a friend	-	-	-	-
9	Ability to perform tasks	38	34	34	46
10	Lack of numerical numbers	-	2	-	-
	Reasons for NOT being confident				
1	Problems with language	28	14	6	
2	Problems due to complex tasks			32	10
3	No interest in the topic		4		
4	Lack of familiarity with the topic	2	2	2	2
5	Lack of planning time	-	4		

As can be seen in Table 21 above the participants' confidence seemed to be related to their self-belief about their ability to perform the tasks and again this was most strongly shown in relation to Task 4 (46%), followed by Task 1 (38%), and then Tasks 2 and 3 (each 34%), e.g.,

Yes I think I was successful because we can explain about the specifications of the comparison between two Blackberries (Task 1).

Once more familiarity with the topic was a key factor contributing to the participants' confidence - Task 4 (26%), followed by Tasks 1, 2, and 3 (24%, 22%, and 10% respectively). Thus the results indicate that the more familiar the participants are with the topic of the task, the more confident they are in doing the tasks to be, regardless of the degree of task difficulty:

I think we did it well because this is a familiar topic for us to make conversation (Task 3).

I feel we did it successfully. I don't know why. It is just like daily conversation so it is not really difficult for us to make conversation so I think I did this task well (Task 3).

Again this is in line with the study by Nunan and Keobke (1995) who suggested that the learners found it difficult to perform the tasks due to three main factors, one of which is learners' lack of familiarity with the topic. It also reflects those factors included in Robinson's (2001, a, b, 2003, 2005) Triadic Componential Framework – namely that factors other than those predicted by way of the Cognition Hypothesis will have a role to play in learners' performance of task.

Some participants also said that they were confident about performing the tasks because of the amount of information that was provided, specifically for two simple and one complex tasks (Tasks 1, 2, and 4). None mentioned this aspect with regard to Task 3 at all. This finding may suggest that tasks with adequate information will improve the participants' confidence in performing the tasks, regardless of the level of task difficulty (cognitive factors). The information may have to do with the instructions or procedures of how to perform the tasks and the detailed descriptions of the topic for discussion.

I think I did the task successfully to compare between these mobile phones because this task talks about the simple thing that we have already known, mobile phones, and also the features that we know have already been included in this task, such as the price, color, and the features are complete (Task 1).

Once more the participants described how the provision of pictures (there were two pictures of Blackberries mobile phones and Houses for lease) contributed to their confidence in performing the tasks, but this time only two for the simple ones (Tasks 1 - 2% and 2 - 4%).

I feel confident because with the picture we can explore again in our brain, in our mind, not like just writing, not just reading, not like just listening. We can explore again more than it (Task 2).

A few other participants were confident about performing the tasks, especially Tasks 2 and 3 (8% and 2%), due to their “interest in the topic” rather than in the manipulation of task complexity, particularly regarding the planning time and the number of elements. In other words, as the participants are interested in the topic of the tasks, they are more confident in performing the tasks, under either simple or complex conditions, e.g.,

I think it's successful because the topic is interesting for me and then I'm more relaxed than the first task I don't know maybe because what is it we're speaking English for long enough duration I think (Task 2).

Similarly, familiarity with performing the previous tasks helped them to feel more confident about performing the task well, although this was only mentioned in relation to Task 3, e.g.,

I feel successful in this task because I can make more conversation in this task better than before because I have already done two similar tasks (Task 3 A2).

Others reported that they felt they could perform the complex tasks well (Tasks 3 - 10% and Task 4 - 12%) e.g.,

It's quite well and I think as the items are a lot so we have many choices and we have many things for us to discuss (Task 3).

One participant described how the lack of numerical figures, especially for Task 2 (2%) helped increase his confidence.

I think it's successful because from the task we can compare and we can make conversation better than the first task because in the first task there are so many numbers and in this task it's just words (Task 2).

The findings suggest that using pictures, repeating tasks, and negotiating topics of interests with learners are worth considering when designing tasks. If this is done it can

have a positive effect on learners' confidence when performing tasks, which can ultimately lead to learners' success of learning English.

Several participants expressed a lack of confidence about performing the tasks for the same reasons that others felt their confidence was improved. Some described how their confidence was diminished because of problems associated with the language of the tasks, the complexity of the tasks, their lack of familiarity with the topic, having no planning time, and their lack of interest in the topic. They said that they could not perform the tasks well because they were not familiar with the topic of discussion for all four levels of tasks e.g.,

I felt unsuccessful maybe because I don't have background and this task has six types and then it made me confused (Task 3).

Participants mentioned that problems with the language inherent in the tasks diminished their confidence in relation to Tasks 1, 2, and 3, (28%, 14%, and 6% respectively), but none commented on this aspect for Task 4. So even though Tasks 1 and 2 are the simplest, the participants felt less confident when performing these tasks due to the language required. Again, it seems that learners' confidence and the problems associated with language are not closely related to the manipulation of task complexity factors (i.e., planning time and the number of elements).

Not really well I think because I'm still too much quiet and it's difficult to answer the dialogue maybe only the vocabulary. I don't really master the vocabulary, my vocabulary is not adequate to make the dialogue. I'm questioning in my mind but I can't say it (Task 1).

I feel unsuccessful because in our conversation we speak in pause too long to think about the words that we want to talk (Task 2).

Participants' lack of confidence due to language problems is consistent with the study by Tavakoli (2009) who found that linguistic demand is considered to be one of the aspects

that leads to more difficulty in performing tasks which may then lead the participants to feel less confident when performing the tasks.

A lack of interest in the topic – another learner factor – was also reported by a few participants (4%) as negatively affecting their confidence in performing the task.

Together these views indicate that the degree of confidence does not closely correspond to the level of task complexity. That is, the participants' degree of confidence in performing the tasks does not necessarily rely on the increasing or decreasing task complexity. The majority of the participants said that they were confident and successful in their performance of the four levels of task, and contrary to what is predicted by the Cognition Hypothesis, the participants were more confident about performing the more difficult planned complex task than the unplanned complex one. This was because their confidence was related to their familiarity with the topic and their familiarity with the previous task, particularly for Task 4, along with a host of other factors as described above.

However, it should be noted that there were a few participants who indicated they felt confident when performing the tasks because there were “many things to compare”. This finding suggests that the increasing complexity of the cognitive factors does not necessarily decrease the participants' confidence in performing the tasks. Again, these findings are in line with Robinson's (2001b) argument that task complexity (a cognitive factor) and task difficulty (a learner factor) do not always have a fixed relationship to each other due to “inherent ability differentials between learners” (e.g., confidence).

5.2.4 Participants' Interest

Almost all of the participants reported that the four tasks were interesting. The one described most in this way was Task 1 (98%), followed by Tasks 3, 2, and 4 (88.46%, 86%, and 82.70% respectively). The participants' reasons for the tasks being interesting to perform can be seen in Table 22 below.

Table 22: Participants' interest in the four levels of tasks

No	Perceptions	Percentage			
	Responses for interest in doing tasks	Task 1	Task 2	Task 3	Task 4
1	Simple task	2	4	-	-
2	Complex task	-	-	12	13
3	Familiarity with the topic	23	19	10	15
4	Interest in the task	13	12	8	10
5	Amount of Information provided into task	6	4	2	2
6	Pictures	13	10	12	6
7	Familiarity with the previous tasks	-	2	-	-
8	Conversation with a friend	2	2	-	2
9	Ability to perform tasks	-	-	2	4
10	Source of motivation in learning English	6	8	17	4
11	A medium to increase speaking in English	15	13	8	8
12	A medium to get knowledge	17	15	19	17
	Responses for NO interest in tasks				
1	Problems due to complex tasks	-	-	4	10
2	Not interesting topics	2	6	8	6
3	Lack of familiarity with the topic	-	6	-	4

As can be seen from Table 22, the participants gave a range of reasons for their interest such as the familiarity with the topic, interest in the topic, amount of information and the pictures, and they did so for all four tasks. They also described their interest in relation to how they believed the tasks improved their capability of speaking English, how the tasks contributed to improving their motivation for learning English, and how the structure of the tasks helped to develop their knowledge of the topic. Other reasons for their interest that were mentioned for some, but not all tasks, included that they were either simple or complex, that their interest was piqued because they had done similar previous tasks, and because their interest aligned with their capability of performing the tasks. Therefore, it would seem that participants' interest in performing the four tasks was due predominantly to learner factors, although manipulation of cognitive aspects did contribute to participant interest, too.

I think this discussion is very interesting because it will make us better in speaking in English. I think like that (Task 1).

I think it's interesting because there are pictures in the task and if we look at the pictures we can imagine the Blackberry in reality. I think that's interesting. That's good (Task 1).

It's very interesting because it compares two houses that each house has more different features but we can choose the best one from our type (Task 2).

On the other hand, some of the participants reported that the tasks were not interesting. They said this was because they lacked interest in, or familiarity with, the topic:

I am not interested in the first topic. It cannot make me interested. It's boring topic and I think and clue information is similar information because the topic is not up-to-date (Task 4).

and, in particular, some did not find the complex tasks interesting.

The information is too long, too complex, and the topic is not up-to-date (T4).

The topic is not interesting, so I could not make dialogue well (T2).

I think it's not interesting because it is about houses and too many houses too rent (Task 4).

In short, the results suggest that the four tasks were, in the main, reported as interesting by the majority of the participants, regardless of the levels of task difficulty. Only a few of the participants were not interested in tasks because the topic was not interesting, they were less familiar with the topic, or the task was too complex (greater cognitive engagement).

The results of this study provide little evidence to indicate a close relationship between the participants' degree of confidence and the four levels of dialogic task complexity (cognitive factors). Almost all of the participants were interested in all four tasks as evidenced by the high number of responses in this regard. Interestingly, the complex task (Task 3), which is predicted by the Cognition Hypothesis to be more difficult than

Task 2, was perceived as an interesting task more frequently than the simple task (Task 2). The participants' interest in all the tasks may indicate that even when tasks are predicted to be complex, as suggested by the Cognition Hypothesis, that is due to the increasing complexity of cognitive factors, is not impossible for learners to perform them if they are interesting.

Tasks 1 and 3 were more frequently mentioned as interesting tasks compared to Tasks 2 and 4. This might simply be because the topics of Tasks 1 and 3 (i.e., Blackberry mobile phones) are considered to be more interesting than Houses for Rent. Despite topics Tasks 1 and 3 being similar to each other, the participants gave different emphasis to the various aspects for the two levels of tasks: Task 1 was regarded as interesting due predominantly to the participants' familiarity with the topic of the task, while Task 3 was deemed interesting due to it enabling them to gain knowledge of the topic. This means that even though the participants are less familiar with the topic of the tasks, they may remain interested in the tasks when the topics enable them to gain new knowledge that they see as useful or relevant.

Furthermore, the participants' perspectives about their interest in the tasks appeared to share a number of common underlying aspects including: familiarity with the topic, interest in the topic, amount of information provided, the pictures provided, that they felt performing them improved their speaking capability, that the tasks improved their motivation for learning English, and the tasks helped develop their knowledge of the topic. Several participants mentioned certain aspects with respect to different versions of tasks. For example, they were interested in performing the tasks because of "having dialogue with a friend" particularly for two simple tasks, and one complex task (Task 4), while none of them commented on this aspect for another complex task (Task 3). Similarly, the task was regarded as interesting due to familiarity with the previous task by only one participant for Task 1 and Task 4. A few participants commented on the aspects of task complexity (cognitive factors) underlying their interest in the tasks, that is, they were interesting due to having "two things to compare" and "many things to compare", which was in reference to simple tasks and complex tasks, respectively. The participants' varied views about the four different tasks suggest that individuals have

different priorities when it comes to their interest in the tasks, some of which lie within the Triadic Componential Framework (i.e., cognitive, interactive, or learner factors) or outside it (i.e., other aspects).

There were a very few participants who perceived the tasks as not being interesting. In the main this was due to three aspects - lack of interest in the topic, lack of familiarity with the topic, and having many things to compare, all of which reflect both the complexity of the tasks (i.e., cognitive factors), but also learner factors (i.e., affective variables). For example, lack of familiarity with the topic was one of the aspects underlying their lack of interest in the tasks. In short, these findings indicate that the participants' degree of interest in the tasks predominantly relies not only on the task complexity (cognitive factors), specifically manipulated within planning time and the number of elements, but also other aspects including interactive and task conditions, and learner factors. Together this reflects both Robinson's Triadic Componential Framework and Skehan's (1996, 1998) concept of task difficulty.

5.2.5 Participants' Motivation

Almost all of the participants said they were motivated to perform the four tasks. Task 1 generated the highest percentage of agreement regarding motivation (100%), followed by Tasks 2, 4, and 3 (94%, 84%, and 77%). Given the similarity of responses, particularly for Tasks 1, 2, and 3, these results suggest that the tasks were motivating for the participants, regardless of the levels of task difficulty. As such it appears that the participants' motivation does not closely correspond to manipulation of the four levels of task difficulty.

As can be seen in Table 23, once again the participants gave a variety of reasons for their motivation including with respect to the tasks themselves how they helped improve their ability in speaking English, how they helped improve their overall level of motivation for learning English, how they enabled them to develop knowledge, that they provided useful information, and that both the pictures and having a dialogue with a friend were motivating.

Table 23: Students' perceptions about tasks and their motivation

No	Perceptions	Percentage			
	Reasons for motivation in performing tasks	Task 1	Task 2	Task 3	Task 4
1	Complex task	-	-	-	2
2	Familiarity with the topic	4	4	2	2
3	Pictures	8	8	2	2
4	Conversation with a friend	4	4	6	4
5	Motivation of learning English	18	32	26	20
6	Improving speaking English	48	30	28	34
7	Getting knowledge of the topic	18	12	18	16
	Reasons for NO motivation in performing tasks				
1	Problems with language	-	2	-	4
2	Problems due to complex tasks	-	-	18	12
3	No interest in the topic	-	2	-	4
4	Lack of familiarity with the topic	-	4	-	-

The reasons varied according to the different tasks. For example, Task 1 was most frequently commented on as being motivating because it improved participants' capacity to speak English (48%).

Yeah I think I would do this task again because it will be helping us to increase our speaking skills and it will be very helpful for us to practice English (Task 1).

Yes I like this activity because this can make my ability in speaking better than before because I will give opinions about the Blackberry, between two Blackberries and I think it can make my speaking better (Task 1).

Similarly Task 1 (18%) was described as motivating because of the knowledge the participants gained from doing it:

Yes if someone asked me to do this task again I want to do it. I think this task is very interesting because we can know things I don't know before (Task 1).

Yes not only we are going to make debate why this is the good one but we also get the information. Blackberry has

the features like this the ringtone is like this the memory capacity is like this. So we not only make a good argument how to be in our opinion but also we get information, get knowledge from this conversation (Task 3).

In contrast, most agreement for Task 2 (32%) was with regard to its contribution for improving English speaking skills:

I want to do it again because it will develop our imagination and our ways to think about how to speak more relaxed and freely and more natural and it's good for studying about grammar (Task 2).

Less frequently mentioned was the motivation that came from having a dialogue with a friend: Task 3 (6%), Tasks 1 and 4 (each 4%), and Task 2 (2%):

Yes I want to do the task again, as I said before this task can improve my speaking in English because we practice speaking English to each other, not in our mind only (Task 3).

In a similar way, the use of pictures as being motivating was mentioned only infrequently for Tasks 1 and 2 (8%) and Tasks 3 and 4 (2%).

Yes, because with the picture we can what is it we can develop our vocabulary because in the picture what is it we can say more because in the picture we can say something (Task 1).

On the other hand, there were a few participants who described how they did not find the tasks motivating. For some this was because the tasks were complex with too many elements to compare, particularly with respect to Tasks 3 and 4 (18% and 12%):

No, I think that's enough because the pictures are too many I think and I have difficulty to compare many pictures like this but if the pictures are only two or three I wanna do (Task 3).

They also described how when they were not interested in the topic, their problems associated with the language of the tasks diminished their motivation, as did their lack of familiarity with the topic. Therefore, it does seem that increasing the difficulty of the tasks (more cognitively demanding) meant that a few participants were less motivated to perform the tasks, although for others the opposite was found.

Hence the findings of this study suggest that there is not a close relationship between task complexity, manipulated simultaneously within the planning time or the number of elements, and the participants' degree of motivation when performing the tasks. This is because almost all of the participants perceived the four tasks as motivating, regardless of level of complexity. These findings support Robinson's (2001b) argument that complexity and difficulty do not always have a fixed relationship to each other for two reasons. First, this is as "a result of inherent ability differentials between learners, that is, differences in the limits of the attentional, memory, and reasoning resource pools". Second, the learners' "inherent ability differentials can also be affected by such temporally limiting factors as motivation" (p.31).

This means that learner factors, task conditions and interactive factors, rather than cognitive factors, led the participants to be highly motivated when performing tasks. Their motivation with respect to learner factors is related to their perceptions that the tasks improve their capacity to speak English, increase their motivation for learning English, and help them acquire knowledge, while task conditions and interactive factors are related to such aspects as, amount of information provided about the tasks, having a dialogue with a friend and the input available from the tasks (e.g., pictures). It appears that the manipulation of task complexity (cognitive factors) within the planning time and the number of elements does not significantly affect the participants' degree of motivation in performing the dialogic task complexity. That is, increasing task difficulty (cognitive factors) within both the resource-directing (the number of elements) and the resource-dispersing dimension (planning time) does not necessarily decrease motivation.

However, there were a few individual who appear as exceptions to this. A very small number of the participants ($n = 4$) said that they were less motivated to perform the tasks

due to there being many elements to compare (complex tasks), because of their lack of familiarity with the topic, lack of interest in the topic, and problems associated with language especially for Tasks 2 and 4. Their lack of motivation in performing the tasks due to their complexity consistently occurred with two versions of complex tasks (Tasks 3 and 4), while lack of interest in the topic and problems associated with the language required for the tasks were commented on for Tasks 2 and 4 (Houses for Rent). A few participants said that they were not motivated when performing the task because of their lack of familiarity with the topic, especially for Task 2, that is, Houses for rent. Thus it appears that for some participants their degree of motivation when performing the tasks maybe affected by cognitive factors. Even so, it would seem that task conditions contributed most to the participants' level of motivation.

5.2.6 Learning opportunities

All of the participants agreed that the two simple tasks (Tasks 1 and 2) provided learning opportunities and the vast majority indicated similarly for Tasks 3 and 4 (94% and 96%). At the same time, the participants more frequently indicated that the two complex tasks (Tasks 3 and 4) were better for improving their spoken language of English (42% and 51%) compared to the two simple tasks (Tasks 1 and 2) (24% and 33%). Even so, the perception of the participants with respect to all four versions of the tasks is that they provided learning opportunities, regardless of the degree of difficulty (cognitive factors) as can be seen in Table 24.

From the participants' comments, there appeared to be a clear link between the perception about learning opportunities and those aspects described previously with regard to motivation. For example:

Yes, because this kind of task gives such kind of incentive especially in my brain to always think and always produce proper sentences to describe something (Task 2).

Table 24: Participants' perceptions about tasks and their learning opportunities

No	Perceptions	Percentage			
	Reasons for providing learning opportunities	Task 1	Task 2	Task 3	Task 4
1	Amount of information provided into task	2	4	-	-
2	Pictures	11	11	4	7
3	Conversation with a friend	4	8	2	-
4	Motivation of learning English	51	33	40	40
5	Improving speaking English	24	33	42	52
6	Getting knowledge of the topic	6	12	6	2
	Reasons for NOT providing learning opportunities				
1	Complex tasks	-	-	2	-
2	Numerical numbers	-	-	2	-

Further, a number expressed a belief that learning opportunities occurred because they could develop their knowledge by performing the tasks.

Yes, we can learn to rent a house for our life later when we are adults or when maybe we are already married so I think it's very good experience to talk about this even though I don't really like it (Task 2).

A few participants also said that the tasks facilitated learning opportunities when they performed them interactively, that is, having dialogue with a friend.

Yes, and I think it gives chance to learn English because in this task we are speaking in English with our friends with other people so we can speak English in our practice (Task 2).

The participants also mentioned, although less frequently, that the pictures that were provided with the tasks facilitated their learning -Tasks 1 and 2 (11%) and Tasks 3 and 4 (4% and 6% respectively).

In contrast to these responses, there were participants who suggested that some aspects of the tasks did not facilitate learning. For example, some suggested having too many elements to compare in the complex tasks diminished their learning. Therefore,

increasing difficulty within the resource-directing dimension (more cognitive engagement) led some to believe that there were less learning opportunities. However, this was the exception. In general, it appeared that the participants did perceive the tasks increased their learning opportunities, regardless of the degree of task difficulty. They indicated that the tasks boosted their capability for speaking, their knowledge and their motivation for learning English.

The findings of this study suggest that there is only minimal evidence to indicate a relationship between the participants' views about learning opportunities and level of task complexity. Instead the findings show that almost all of the participants reported that each task provided opportunities for learning. In addition, the participants predominantly commented on interactive factors (i.e., dialogue with a friend), learner factors (e.g., getting knowledge of the topic) and other aspects, such as pictures, as increasing their opportunities for learning English and improving their capability of speaking English, rather than reflecting on task complexity as the source of learning.

However, one pattern did emerge in relation to this: The tasks were considered to provide learning opportunities if the participants were encouraged to have dialogue with a friend, especially for the three levels of task complexity, but not for the unplanned complex task (Task 4). This indicates that working with a partner in conversation is perceived by learners as providing learning opportunities, but that in their perception this was mediated to a certain degree if the task was considered to be too difficult. .

Another aspect perceived as providing learning opportunities is the amount of information given. For example, the participants described how the information helped their learning:

The task included complete information of the houses so it can encourage me to speak more fluently (Task 2).

According to Skehan (1998, p. 99), this aspect (i.e., the amount of information) related to “cognitive processing” which includes “information organization, clarity and sufficiency of information given, and information type”.

As indicated, in contrast to the findings that learner and task conditions are perceived to provide learning opportunities, increasing the complexity of the tasks (i.e., cognitive factors) was commented on less frequently. Even so, the two complex tasks which, as predicted by the Cognition Hypothesis should be more difficult than the simple ones, were more frequently commented on as being able to provide learning opportunities compared to the two simple tasks, that is, improving capability of speaking. Therefore, even though the tasks are regarded as cognitively more demanding, in the learners' perceptions they may provide learning opportunities if they encourage the participants to speak English, if it has pictures, or the topics provide new information or knowledge.

Thus, it seems that whilst aspects of task complexity, especially planning time and the number of elements, may provide learning opportunities, so too do other factors. However, in terms of pedagogy, these other factors, as Robinson (2001a, b, 2003, 2005) argues, are difficult to predict in advance.

In short, in this study there appears to be only a minimal relationship between the manipulation of task complexity and the participants' views about learning opportunities. These findings suggest that in the perception of the learners, interactive and learner factors contribute to opportunities for learning.

5.2.7 Dialogic Nature of the Tasks

All of the participants commented on the benefits of performing the dialogic tasks. Their comments regarding the four levels of dialogic features of tasks can be seen in Table 25 below.

Table 25: Participants' perceptions about dialogic features of the four tasks

No	Perceptions	Percentage			
	Advantages of dialogic features of tasks	Task 1	Task 2	Task 3	Task 4
1	Improving speaking English	33			
2	Facilitating with peer corrections	60			
3	Feeling free to express ideas	7			

Generally, the participants spoke positively about the four tasks. They particularly commented on their enjoyment and the benefits of performing them. For example, they described how when performing the tasks they felt relaxed and confident, received useful peer correction (60%), improved their capability of speaking English (33%), felt motivated, and they enjoyed feeling free to express their own ideas (7%):

Yes because we can share with my friend if we don't understand we can ask my friend so we can help each other to do the task (Task 4).

Yes I like this task because if I discuss with a friend I am not shy to express ideas and I can ask my friend when I make errors (Task 4).

Yes because this is very good for daily conversation because the condition can improve our skill in speaking and I think I would try it in my boarding house with my friend (Task 4).

These views suggest that having dialogue with a friend and working interactively, regardless of the degree of task complexity, leads the participants to believe that they benefit from the experience. Further, they enjoy working interactively in pairs. They also feel it is beneficial to get feedback from their partners when working collaboratively and appreciate the help they get from their partner when communication breaks down , believing this facilitates their improved performance of spoken language.

Some participants also said that they felt a sense of freedom when discussing parts of the tasks with friends. They were no longer “shy to express ideas”. In turn, this situation may enable learners to increase their motivation which also contributes to their success in learning English.

Hence, these findings suggest that interactive factors (i.e., dialogic features of tasks) may play an important role in facilitating the learning of English. This is in line with Long's (1996) argument that second language learners need opportunities for

meaningful interaction in order to assist their acquisition, and that such opportunities can be created through the use of tasks and task based teaching approaches.

Furthermore, participants' views about dialogic (interactive) tasks seem to confirm those advantages outlined in the literature (e.g., Anton, 1999; Ellis, 2003; Robinson, 2007). According to Anton (1999), "through dialogic interaction, teachers can provide learners with effective assistance that will enable them to perform at higher levels than they would otherwise" (p. 304). Ellis (2003) makes the point that "dialogic discourse serves to create the intersubjectivity that enables verbal interaction to mediate learning" (p. 177). Robinson (2007) further argues that interaction is "an important context and opportunity for activating process thought to contribute SLA" (p. 14). In short, when learners are provided with opportunities to engage interactively by working in pairs or groups, this facilitate their language learning.

5.3 Summary

There was some evidence indicating that the participants' perceptions reflected the level of task difficulty as manipulated according to planning time and number of elements (i.e., the cognitive factors). For instance, a number of the participants found the complex tasks (increasing cognitive difficulty) more difficult and stressful. However, other factors contributed to the participants' perceptions of the tasks. For example, familiarity (of topic and task procedures), their confidence, interest, motivation, and the dialogic (interactive) nature of tasks also had an impact. In this way the participants' perceptions did not completely align with the four levels of task difficulty. Even so, the four tasks did enable the participants to be actively involved in learning English. The participants' perceptions do highlight the importance of all the factors included in Robinson's Triadic Componential Framework (i.e., cognitive, interactive, and learner factors), but also other factors, such as the contribution of input (e.g., pictures).

Chapter Six

FINDINGS: L1 RATERS' JUDGEMENTS ON INDONESIAN PARTICIPANTS' PERFORMANCE

6.1 Overview

This Chapter addresses Research Question 3 and reports on the findings of L1 Australian English raters' judgments of the oral production by the Indonesian participants. It includes both numerical ratings and written comments. These data were used to investigate the degree of fit between L1 raters' judgments and CAF with respect to the oral production of the Indonesian participants.

6.2 L1 Australian English Raters' Assessments of the Oral Production of the Indonesian Participants

The following table presents the results of the three L1 raters on the Likert scales.

Table 26: Three L1 raters on Likert scales for the four levels of tasks

L1 raters	Task 1		Task 2		Task 3		Task 4	
	Ranges of ratings	\bar{A}	Ranges of ratings	\bar{A}	Ranges of ratings	\bar{A}	Ranges of ratings	\bar{A}
L1 A	2 - 5	3.2	2 - 6	3.1	2 - 5	2.9	2 - 4	2.7
L1 B	4 - 6	4.8	4 - 6	4.7	3 - 6	4.7	4 - 5	4.6
L1 C	2 - 7	5.5	3 - 7	5.2	3 - 7	5.1	2 - 7	5.0

\bar{A} : rating average

As can be seen in Table 26 above, the L1 raters' judgments of the Indonesian participants' oral production, there was a high degree of agreement across the three raters for the four levels of task difficulty, although there was some variation. A summary of the level of agreement is shown in Table 27 below. From this it can be seen

that Task 2 had the highest percentage of agreement (88%), followed by Tasks 3 and 4 (86%), and Task 1 had the lowest level (80%).

Table 27: Rating agreements of three L1 raters on oral production for four levels of tasks

No	Criteria of agreement	Task 1	Task 2	Task 3	Task 4
		%	%	%	%
1	Three ratings are the same.	2	6	2	0
2	Two ratings are the same and the third within one or two points (Likert scale).	34	38	29	27
3	Three ratings within one point (Likert scale).	19	21	34	27
4	The two ratings are within one point and the third rating is within two points (Likert scale).	25	23	21	33
Total		80	88	86	86

Further the L1 raters' assessments and comments suggest that they considered, "the Indonesian learners of English have approximately the same levels of proficiency". This is in contrast to the CAF results which showed considerable variability between the Indonesian participants suggesting that their level of proficiency was not the same.

With respect to the written comments, it appeared that the three L1 raters generally assessed the oral production of the Indonesian participants according to four criteria - pronunciation, fluency, linguistic features (e.g., syntax, tense), and extra-linguistic categories, as outlined in Table 28 below. Overall, there were 180 comments for these four issues with pronunciation most frequently mentioned, followed by fluency, linguistic features, and then extra-linguistic categories.

Similar to the results of the rating scales, there was broad agreement between the three L1 raters' written comments. It is interesting to note that this occurred even though the three raters approached the judging task in different ways. Raters A and B commented globally on the oral production for each individual while Rater C judged the oral production of each Indonesian participant on all four tasks.

Table 28: L1 raters' written comments on the Indonesian oral production for four versions of tasks

No	L1 raters' Assessment		Frequency of comments	Percentage of total comments
1	Pronunciation	General Pronunciation (i.e., mispronounce of single sounds and words)	78	44
		Accent, Stress, Intonation	7	4
2	Fluency	Speech flows/fluidity (smoothness)	28	16
		Hesitation and pauses	7	4
		Speech rate	7	4
		General fluency (non-specific)	6	3
3	Linguistic Features	Syntax	15	8
		Vocabulary	8	4.4
		Omission (e.g., the copular verbs, articles)	10	5.5
		Comparative Degree	7	4
		Verb Tenses	3	1.6
4	Extra-linguistic aspects		5	2
Total			180	100

Although the three raters broadly agreed in terms of the rating scales and in their written comments, their individual judgments of the oral production of the Indonesian participants differed. However, given the design of this study (i.e., the use of only three raters), the data of L1 raters should be understood to be indicative rather than definitive.

Rater A

In general, Rater A rated the Indonesian participants' oral production lower compared to the other two raters. Her ratings ranged from the lowest (2) to the highest (6) within the ten-point Likert scale. Her ratings across all four tasks are similar, although, her ratings for Task 2 (the unplanned simple task) were, on occasion, slightly higher. With respect to her written comments, Rater A's comments were about all of the Indonesian

participants' oral production rather than about individual performances. In the main her comments focused on three issues - pronunciation, fluency, and syntax. Firstly, her comments about pronunciation problems centered on the participants mispronouncing certain words or sounds.

The main point being some trouble with pronunciation.

Some words were more difficult to understand than others.

The improvements could be made by speaking at the same cadence.

Her comments on fluency appeared to focus on speech flow or 'smoothness' and speech rate, suggesting that this rater viewed fluency from a broad perspective.

Although I could understand the English spoken but found it lack of fluidity.

Speech alternated between fast and slow.

With respect to syntax her comments were often of the following kind.

Sentence structure was confusing at times.

Rater B

Rater B scored the Indonesian participants higher compared to Rater A. He rated their production from the lowest score of 3 to the highest of 6. Unlike Rater A, he tended to provide the lowest rating (3) for the complex tasks, Tasks 3 and 4, although generally his ratings were quite similar for all four tasks. Like Rater A, with respect to the written comments, Rater B commented globally, in summary form, on the oral production of all 52 of the Indonesian participants, for example, he stated that in his judgment the Indonesian participants were of the same level of proficiency.

He also commented on four criteria in his written comments - pronunciation, fluency, linguistic features, and extra-linguistic aspects. His comments on pronunciation were

more general, rather than being about specific features of pronunciation. This is demonstrated in his comment

Some students had minor problems with pronunciation.

Similarly his comments about fluency were quite general.

Some students were hesitant with their replies when asked a question which affected fluency of the conversation.

With respect to linguistic features, he emphasized mostly those problems associated with vocabulary.

The vocabulary of all students was very good but limited.

Unlike Rater A, Rater B also commented on the extra-linguistic aspects of the participants' speech.

Some students sounded a little more confident than others.

Rater C

In general, Rater C gave higher ratings for the participants' oral production on all four tasks compared to the other two raters. She also provided more varied ratings compared to Raters A and B. Her ratings ranged from the lowest rating of 2 to the highest of 7 within the ten-point Likert scale. Her lowest scores were mostly generated from Task 1 (the planned simple task), although she was generally consistent across all four tasks.

Unlike Raters A and B, Rater C commented on the oral production generated by every individual Indonesian participant as they performed all four tasks. However, like Rater B, Rater C commented on four issues in her written comments, pronunciation, fluency, linguistic features, and extra-linguistic aspects. Overall, she commented on the features of pronunciation more frequently than the three other issues. Pronunciation was often

referred to in a general sense, or as mispronunciation of words or single sounds. Rater C also commented on specific features of pronunciation, such as intonation or accents.

Some pronunciation needs improvement.

Pronunciation of many words makes listening a little tricky.

Accents on syllables sometimes wrong, making hard to grasp.

Like Rater B, Rater C also commented on the extra-linguistic aspects in her judgments, but she appeared to consider more features within this category compared to Rater B.

Shows emotion.

Good grasp of conversation

Personality comes through.

More confident speaker

Rater C also commented on linguistic features, particularly issues relating to syntax, omission, vocabulary, comparative degree, and verb tenses.

Leaves plurals of some words, struggle to translate.

Speaks well but stumbles on some words.

Grammar is funny at times, e.g., more cheaper.

Verb tenses sometimes wrong, e.g., have instead of has.

Furthermore, Rater C commented in detail on fluency in her written comments. She considered four features of fluency, speech flow (smoothness), hesitation (pauses), speech rate, and general fluency. Speech flow was the most frequently mentioned criteria of fluency, while hesitation or pauses and speech rate were the second highest (each seven times). General fluency was the least frequently mentioned criteria (six times).

Speech flows better

Lack of fluidity.

Pauses a lot

Speaker B speaks very quickly

Speaker A speaks more fluently than Speaker B.

In summary, the three L1 raters' judgments on the Indonesian participants' oral production were largely in agreement in terms of rating scales and written comments. With respect to their written comments, their perspectives appeared to be similar in terms of pronunciation, fluency, and linguistic features. However, only two of the Raters (B and C) appeared to consider the extra-linguistic aspects of the participants' production. With respect to pronunciation, each commented slightly differently. For example, Raters A and B commented on pronunciation in general terms, while Rater C commented on a greater variety of features, including intonation and stress.

With respect to fluency, Rater A commented on features of speech flow and speech rate, whilst Rater B paid attention to hesitations. Rater C again commented on a greater variety of fluency issues compared to Raters A and B. All three raters commented on linguistic features, but there were differences and similarities with respect to these. Rater A commented on syntax, while Rater B emphasized vocabulary, and Rater C mentioned both syntax and vocabulary and also other linguistic features (e.g., omission, comparative degree, and tenses).

Overall, however, the three L1 raters appeared to have similar impressions of the Indonesian participants' oral production. Furthermore, the four different levels of task difficulty did not overly impact on their judgments.

The results from the three L1 raters were also compared with CAF scores generated from the participants' performance of four tasks. As described in Chapter Three, the degree of fit between L1 raters' judgments and CAF were examined by comparing the trends or patterns of L1 ratings and CAF measures and the L1 raters' written comments were used to elaborate their ratings. The comparison of L1 raters' ratings and written comments and CAF are reported and discussed in the following sections.

6.2.1 Comparisons of L1 Raters' Judgments and Complexity Measures

The data of L1 ratings and Complexity measures are presented in Table 29 below. The results show that there is not a close alignment between the scores of the three L1 raters and the Complexity measures across the four tasks., that is, high ratings from the L1 speakers did not necessarily mean higher scores on complexity measures for syntactic constructions, lexical words, and Guiraud's Index and vice versa. For example, for Task 1 when a rating of 7 was given by the three L1 raters, the range for complexity scores ranged from 1.3 (lowest) to 1.6 (highest) for Syntactic Complexity. For Task 1, the three L1 raters gave lowest rating, namely 2, whereas the scores on Syntactic Complexity measures ranged from 1.4 to 1.9. Similar patterns between L1 ratings and Syntactic Complexity measures also occurred with the other three versions of the tasks.

Therefore, in response to Research Question 3, these findings suggest that there is little evidence to indicate a close fit between the L1 speakers' ratings of the Indonesian participants and the three Complexity measures. This is true in terms of both the rating scales and the written comments provided by the L1 raters. For instance, in their written comments, the three L1 raters commented mostly on the participants' oral production in terms of syntax, omissions, vocabulary, comparative degree, and verb tenses, and did so with respect to the participants' errors, (e.g., ”*Let we compare them.”) rather than on the complexity of their syntactic constructions, that is, the L1 raters' comments did not encompass content that reflected the definitions of complexity in terms of Percentage of Lexical Words or Guiraud's Index of Lexical Richness. Similarly with respect to lexical complexity their comments were mostly about the participants' errors such as the omission of plurals, the copular verb *be*, and wrong verb tenses, (e.g., *have* instead of *has*). Interestingly these findings are consistent with those of previous studies (Kim, 2009; Zhang & Elder, 2010) which suggest that L1 raters do not base their judgments on the type of complexity inherent in such measures.

Table 29: Comparisons of L1 ratings and Complexity measures for four versions of tasks

Rating scales of three L1 raters	Ranges of Complexity measures		
	Syntactic Complexity	Percentage of Lexical Words	Guiraud's Index
Task 1			
7	1.3 – 1.6	12 – 17	6
6	1.2 – 2.1	12 – 25	5 – 8
5	1.0 – 2.0	14 – 27	3 – 8
4	1.1 – 2.0	14 – 25	3 – 8
3	1.1 – 1.8	12 – 27	5 – 7
2	1.4 – 1.9	18 – 22	6
Task 2			
7	1.9	15	7
6	1.3 – 2.2	13 – 21	5 – 7
5	1.2 – 2.2	12 – 30	5 – 7
4	1.2 – 2.0	13 – 30	5 – 7
3	1.3 – 2.1	13 – 27	5 – 7
2	1.2 – 1.8	18 – 27	5 – 7
Task 3			
7	1.3	23	7
6	1.1 – 1.8	10 – 20	5 – 7
5	1.1 – 1.9	10 – 24	4 – 7
4	1.1 – 1.9	12 – 24	4 – 7
3	1.1 – 1.9	12 – 24	4 – 7
2	1.1 – 1.9	13 – 24	5 – 7
Task 4			
7	1.2 – 1.7	12 – 19	6
6	1.1 – 1.8	11 – 21	5 – 7
5	1.1 – 2.0	11 – 29	5 – 7
4	1.2 – 1.8	11 – 20	5 – 7
3	1.1 – 1.8	12 – 29	5 – 7
2	1.1 – 1.7	12 – 29	5 – 7

In addition, the three L1 raters were mainly concerned with the participants' pronunciation, rather than those linguistic features reflected in the complexity measures. They most frequently commented on learners' problems regarding their mispronunciation of words or single sounds, intonation or accents. From their comments it appears that such errors are regarded by the L1 raters as being irritating. It is possible that this is because such errors may result in misunderstanding between a speaker and a

listener and, in turn, this causes difficulties in the conversation. Overall, therefore, mispronunciation can affect “intelligibility”, which may ultimately lead to communication breakdown and so it is understandable that L1 raters focused on this as the important aspect in understanding oral messages. These findings are in line with those found in the previous studies (Fayer & Krasinski, 1987; Jenkins, 2002; Kim, 2009). On this basis Jenkins (2002) suggests that L2 learners need to practice their pronunciation so that communication breakdown can be minimized.

The L1 raters (especially B and C) also commented on extra-linguistic categories rather than just on complexity as reflected in CAF measures (i.e., Syntactic Complexity, Percentage of Lexical Words, and Guiraud’s Index of Lexical Richness). Although these extra-linguistic aspects (e.g., Some students sounded a little more confident than others) were commented upon by the L1 raters less frequently than other aspects (i.e., pronunciation, fluency, and linguistic features), it does suggest that they contribute in important ways to the assessment of oral language. Together it also suggests that L1 raters take into account a broad range of issues when attempting to understand and then to rate oral messages, rather than simply paying attention to linguistic features as reflected in complexity measures.

6.2.2 Comparisons of L1 Raters’ Judgments and Accuracy Measures

The findings of this study indicate that L1 ratings and the three Accuracy measures (i.e., Error-Free AS-Units, Percentage of Self-Repairs, and Ratio of Repaired to Unrepaired Errors) across the four tasks are also dissimilar as can be seen in Table 30 below.

Table 30: Comparisons of L1 ratings and Accuracy measures for four versions of tasks

Rating scales for three L1 raters	Accuracy Measures		
	Error-Free AS- units	Percentage of Self- Repairs	Ratio of Repaired to Unrepaired Errors
	Task 1		
7	48 – 60	0 – 4	0 – 4
6	36 – 82	0 – 40	0 – 67
5	11 – 82	0 – 40	0 – 67
4	31 – 82	0 – 40	0 – 67
3	11 – 65	0 – 40	0 – 67
2	31 – 57	0	0
Task 2			
7	43	3	3
6	33 – 65	0 – 9	0 – 10
5	11 – 82	0 – 22	0 – 29
4	15 – 76	0 – 9	0 – 10
3	8 – 64	0 – 25	0 – 33
2	15 – 66	0 – 43	0 – 75
Task 3			
7	61	10	11
6	44 – 64	0 – 6	0 – 6
5	29 – 68	0 – 13	0 – 14
4	32 – 68	0 – 13	0 – 14
3	29 – 67	0 – 13	0 – 14
2	32 – 65	0	0
Task 4			
7	65	7	8
6	24 – 66	0 – 25	0 – 33
5	19 – 66	0 – 43	0 – 75
4	19 – 66	0 – 26	0 – 35
3	31 – 62	0 – 43	0 – 75
2	19 – 55	0 – 43	0 – 75

For example, whilst the L1 rating of oral production for Task 2 was 6, the scores for Error-Free AS-Units ranged from 33 to 66, and when the L1's rating was 4 the range for Error-Free AS-Units was 15 to 76. A similar wide range of scores and lack of alignment with L1 rater assessments occurred for Percentage of Self-Repairs and the Ratio of Self-

Repaired to Unrepaired Errors across the four levels of tasks. Like the Complexity measures, the findings of this study suggest that there is little evidence to indicate closeness of fit between L1 raters' assessment and the Accuracy measures across the four levels of task difficulty.

However, the L1 raters did comment on the linguistic accuracy, particularly errors relating to syntax (fifteen times), omission (ten times), vocabulary (eight times), comparatives (seven times), and verb tenses (three times). Again this is reflected in the findings of previous studies (Kim, 2009; Zhang Elder, 2010).

From these findings it would seem that what is regarded as 'accurate' as reflected in the Accuracy measures does not closely align with what 'feels accurate' to the L1 raters. It is also possible that the concept of 'accuracy' used by the L1 raters may be broader than the Accuracy measures. They seemed to include other aspects, such as extra-linguistic aspects, as commented upon by Raters B and C (e.g., Personality comes through). Further it seemed that the L1 raters used their intuition in judging the accuracy of the participants' language production, and although these did not align with Accuracy measures, their intuitive judgments, as Davis (2003, 2011) argues, can still be correct. In summary, therefore, the findings of this study suggest that there is no evidence to indicate closeness of fit between L1 raters' judgements in terms of rating scales and Accuracy as measured by Error-Free AS-Units, Percentage of Self-Repairs, and Ratio of Repaired to Unrepaired Errors across the four levels of task difficulty. Further, it seems that L1 raters assess learners' performance in a broader way than is encompassed in the Accuracy measures. However, L1 raters' written comments do reflect Accuracy measures, in the sense that L1 raters also consider linguistic features as important factors in understanding oral language production.

6.2.3 Comparisons of L1 Raters' Judgments and Fluency Measures

As can be seen from Table 31, a comparison of the L1 raters' assessment and the fluency according to Unpruned Speech Rate A and Pruned Speech Rate B once again seemed to indicate irregular patterns of scores across the four versions of tasks, that is, the L1's higher ratings did not always correspond with higher fluency measures and similarly lower ratings did not always align with lower fluency scores.

Table 31: Comparisons of L1 ratings and Fluency Rates A and B for four levels of tasks

Rating Scales for three L1 raters	Fluency Speech Rate A	Fluency Speech Rate B
Task 1		
7	126 – 134	68 – 141
6	16 – 187	73 – 174
5	80 – 223	68 – 192
4	60 – 223	55 – 192
3	60 – 223	55 – 192
2	60 – 103	55 – 95
Task 2		
6	94 – 170	76 – 150
5	73 – 170	57 – 167
4	71 – 162	47 – 151
3	71 – 170	52 – 150
2	71 – 141	52 – 123
Task 3		
7	112	103 – 141
6	104 – 192	100 – 182
5	66 – 192	64 – 182
4	66 – 159	58 – 148
3	77 – 192	58 – 182
2	66 – 150	58 – 156
Task 4		
7	143	137
6	112 – 177	88 – 162
5	78 – 177	59 – 163
4	82 – 177	76 – 162
3	78 – 172	59 – 163
2	78 – 157	59 – 140

For example, the L1 ratings of 5 for Task 1 had a range of Fluency Speech Rate from the lowest (80) to the highest (223). A similar pattern also occurred when the L1 ratings

were compared with the Fluency of Speech Rate B. It is possible, therefore, that L1 raters viewed fluency differently to the way fluency is measured by Unpruned Speech Rate A and Pruned Speech Rate B.

However, unlike the comparison between the L1 ratings and Complexity and Accuracy measures, there did appear to be some concordance between the L1 raters' highest and lowest ratings and the Fluency measures. That is, the highest L1 ratings tended to correspond to the highest fluency scores and similarly the lowest L1 ratings with the lowest fluency scores as measured by both Unpruned Speech Rate A and Pruned Speech Rate B. This finding suggests that the L1 speakers may have rated learners' oral production according to speech rate.

Although this correspondence was not tested statistically because of the nature of the data (i.e., limited number of raters, lack of normalised scores), the written comments of the L1 raters do seem to confirm they did indeed base their judgments on broader issues of fluency. They commented on the participants' fluency in terms of speech flow/smoothness, hesitation/pauses, speech rate, and general fluency. Speech flow was most frequently mentioned, while hesitation or pauses and speech rate were next most frequently listed in the raters' comments, and general fluency the least frequently mentioned criteria. The L1 raters also commented that fluency is the source of 'intelligibility problems'. In this way the findings of this study share similarities with those of previous studies (Fayer & Krasinski, 1987; Kim, 2009; Zhang & Elder, 2010). For example, the L1 raters in this study commented on pausing as did those in the study of Fayer and Krasinski (1987). Fluency in terms of smooth flow of speech and pausing is also in line with the findings of Kim (2009), and in Zhang and Elder (2010).

It does seem that the way the L1 raters assessed oral production in terms of fluency was broader in perspective and they did not simply focus on a speech rate. This suggests that these raters may have prioritized pragmatic fluency, that is, "the combination of both pragmatic appropriateness and smooth continuity in ongoing talk" (House, 1996, p. 228) rather than simply focusing on "speech rate" in their assessments.

6.1 Summary

Generally, the findings of this study suggest that there is not a closeness of fit between CAF and L1 raters' judgments (i.e., ratings and written comments) of the Indonesian participants' oral production. That is, CAF, as a construct to measure oral language production did not correspond closely to L1 raters' judgments. However, there was somewhat of an alignment between L1 fluency ratings and the measurement of fluency according to speech rate. However, the L1 raters seem to prioritize features such as pronunciation and extra-linguistic aspects which are not accounted for by CAF measures.

Chapter Seven

CONCLUSIONS

This study provides evidence that the manipulation of cognitive factors, namely the resource-directing and the resource-dispersing dimensions (i.e., planning time and number of elements), in dialogic tasks can play a role in facilitating learners' oral production in terms of CAF. This evidence does provide some support for Robinson's (1995, 2003, 2005) claim that cognitive factors should be considered in sequencing pedagogical tasks.

At the same time, however, the manipulation of cognitive factors alone was not sufficient to account for learners' performance in terms of CAF. In fact, their performance in this study was not wholly consistent with the prediction of the Cognition Hypothesis. Instead, other interrelated factors such as, *task conditions* (interactive factors), *task difficulty* (learner factors), and *input* (linguistic and non-linguistic sources) also played an important role in performance.

Furthermore, simply judging learners' performance in terms of CAF with a view to sequencing tasks is not sufficient. More holistic measures of learners' performance are also required, including consulting with the students about their perceptions, and judging performance using L1 speakers' assessment.

With respect to learners' perceptions about the four versions of dialogic tasks, generally the more complex the task that is manipulated within cognitive factors (+/- planning time and +/- few elements), the more difficult and stressful learners perceived them to be. However, learners' degree of difficulty and stress in performing tasks was not always due simply to the manipulation of cognitive factors. Moreover, the degree of learners' other affective factors (e.g., confidence, motivation, interest, learning opportunity) in performing tasks was not closely related to the increasing task difficulty of cognitive factors.

In addition, familiarity with the tasks or the effect of repeating similar tasks led the participants to perform tasks more easily, regardless of the levels of task difficulty. The roles of familiarity with the tasks or repetition of performing tasks are important to decrease the feeling of difficulty in performing tasks; therefore, learners can perform the tasks more easily (Skehan, 1998; Bygate, 1999). Repetition or rehearsal does not mean that learners should memorize lists of vocabulary and grammatical rules without a meaningful context. Rather the results suggest that they need to be provided with tasks or activities of the same type or content on a regular basis that encourage them to practice the language in a meaningful situation or context activities in the way language is used as a means of communication as reflected in real-life activities. Therefore, tasks or learning activities can be ‘manipulated’ in such a way that learners are actively involved in undertaking familiar and meaningful activities.

In terms of L1 speakers’ judgments and learner task performance, inconsistencies were found between these and more traditional CAF measures. Firstly, pronunciation is not usually included as a part of assessing learners’ performance in terms of CAF, yet it has been found in a number of studies involving L1 raters to be a source of intelligibility problems (Fayer & Krasinski, 1987; Jenkins, 2002; Kim, 2009; Zhang & Elder, 2010). This current study also found that pronunciation was the most frequent issue commented on by L1 raters, and was the most problematic component in understanding learners’ spoken language. Secondly, CAF excludes non-linguistic expressions, which can be regarded as an important part of the message by L1 speakers in understanding spoken language. Lastly, multiple measures of CAF often resulted in contradictory performance indicators. For example, a learner’s performance in terms of fluency was found to be different when it was measured by Unpruned Speech Rate A and Pruned Speech Rate B. On the one hand, it can be regarded as fluent when measured by Speech Rate B, but not fluent when assessed using Speech Rate A. Amongst researchers there is little consensus about the most appropriate measures amongst the multiple measures of CAF used to assess learners’ performance. Clearly there is a great deal more research to be done in this area.

Despite this, in a broader sense this study does have considerable benefits, particularly for the context in which it was conducted namely English language teaching in Indonesia. It represents a deliberate attempt to shift the paradigm from “synthetic” practices of language teaching, which solely *focus on forms* to task-based approaches. These more contemporary pedagogies have the potential to shift English teaching to a situation where more attention is paid to meaning as well as providing an opportunity to *focus on form*. If successful, this will help address the current shortcomings of teaching English at all education levels in Indonesia which is generally regarded as unsuccessful (Setyadi, 2009; Saragih, 2009; Kasihani, 2010; Zein, 2010). Despite the students having learned English for years, the majority of them still cannot speak in English, even at a very basic level of communication.

More specifically, this study may provide direction to and beneficial outcomes for Indonesian EFL teachers, syllabus designers and students. It provides English teachers and syllabus designers with a framework for designing pedagogical tasks, and factors to consider in their sequence of presentation. Further, such a framework will support them to design a model of appropriate pedagogical tasks that incorporate not only cognitive but also other factors (i.e., learner, interactive and input). This model of pedagogical tasks is expected to facilitate Indonesian EFL learners of English to develop their capability in spoken English. This is because students are provided with communicative activities that may keep them stimulated in their English learning in a more natural way. This is in line with the claim by Long (1991) that “to learn a language is not by treating it as an object of study, but by experiencing it as a medium of communication” (p.41).

This study also contributes to our understanding of task-based learning, especially from an information-processing perspective. By undertaking the study in Indonesia it extends investigations in this area into a new context. It also serves to enhance discussions about the appropriateness of using CAF measures as constructs of proficiency. Specifically, it raises the question of whether relying solely on linguistic considerations, is sufficient. However, much more research is required in this regard particularly within the Indonesian context.

Furthermore, the methodology used in this study, namely a mixed method approach, enabled an in-depth, but also complementary exploration of the data.

This current study has raised a number of issues that need to be addressed in future research. Firstly, since this research included only a small number of participants with the same proficiency levels and mostly of the same gender, further research should include a larger number of participants (for the investigation of the effect of task complexity) by including not only participants from different levels of proficiency but also of different gender. This would enable the comparison of gender in relation to the perceptions of task complexity and the impact of gender in a dialogic task (dyads), for example, the difference of dyads of the same and different gender groups. Since the current study only includes learners with the same proficiency level, future studies might include different levels of proficiency, to investigate the relationship between learners with the same or different levels of proficiency (e.g., low and low, low and high, or high and high proficiency).

Secondly, the findings of the current study suggest that the aspects of the Triadic Componential Framework (i.e., cognitive factors, interactive factors, learner factors), and input (linguistic and non-linguistic sources) contribute significantly to learners' performance. Based on this evidence, task-based studies for future research also need to explore the relationship between these factors, that is, the extent to which the aspects of these interrelated factors contribute to learners' performance in terms of CAF. This follows Robinson's argument that task cognitive factors should be the sole basis for sequencing pedagogical tasks (2001a,b, 2003, 2005, 2007). Moreover, so far studies of task complexity have investigated the manipulation of only certain aspects of task complexity (cognitive factors), either the resource-directing or the recourse-dispersing dimension. This research has provided evidence that even when the task complexity is manipulated within certain aspects (i.e., planning time or the number of elements), other aspects are inevitably incorporated. Therefore, studies to investigate the extent to which each aspect of both the resource-directing (i.e., +/- few elements, +/- here and now, +/- no reasoning demands) and the resource-dispersing dimensions (+/-planning time, +/-

single task, and +/- prior knowledge) contribute to language learners' performance need to be conducted.

Thirdly, the manipulation of planning time in the current study appears to have not shown consistent results in the sense that under the unplanned task conditions, which should have been regarded as more difficult than the planned condition, this was not demonstrated. As previously discussed, familiarity of doing the previous tasks is considered to be one of the aspects underlying the participants' feeling of ease in performing the tasks. Although an attempt was made to minimize repetition effects due to doing previous tasks, it seems that the effect of repetition remained. Therefore, future research, which employs planning time, needs to resolve the overlap of the roles of planning time and task repetition so that clear differences between them can be investigated.

Fourthly, so far, investigating learners' perceptions in terms of task difficulty have primarily paid most attention to their views about performing different levels of the cognitive complexity of tasks. Therefore, further research should be developed to explore the relationship between participants' perceptions of task difficulty and their performance in terms of CAF to investigate whether or not there is much evidence to indicate a relationship or degree of fit between learners' perceptions of task difficulty and their language performance in terms of CAF, that is, whether or not the learners with 'positive' responses of a certain level of tasks perform higher on CAF and vice versa.

Finally, with respect to L1 raters, since there was only a small number (three) of L1 raters participating in this study, further research is suggested to include more L1 raters. This is to investigate whether more L1 raters result in different 'pictures' of assessments of EFL learners' oral production. In addition, the assessment of oral production generated by more L1 raters could then be generalized. Further research could also compare oral production as measured by CAF and L1 raters who have had formal training in linguistics and those who have had no such training. It is an area worthy of future study.

As with most research, although a number of measures were put into place to ensure reliable and valid data, the current study does have three main limitations. Firstly, since the four levels of tasks were administered to the same participants, and even though the role of Speaker 1 and Speaker 2 alternated for the different tasks, it was possible that the participants repeated the same pattern of dialogue and, therefore, the learners' performance may have been affected by this. Furthermore, different topics of tasks between the planned and unplanned conditions might have also affected learners' performance in terms of CAF.

Secondly, since the participants were interviewed in the same pairs as they performed the tasks, they may have imitated each other's responses. In addition, the interviews were conducted in English, as the participants were majoring in English although Indonesian (Bahasa Indonesia) was sometimes used in the feedback loop or for clarification, however, the use of English might have affected their motivation to express more opinions due to their limited ability in speaking in English.

Thirdly, as previously mentioned, there was only a small number (three) of L1 speakers participating in this study. As such, the L1 speakers' judgments of oral production should be regarded as indicative rather than definitive. Therefore, the results of the current study with respect to L1 speakers' judgments of oral production cannot be generalized.

In conclusion, learners' performance in terms of CAF is affected by a number of interrelated factors including those aspects described within the Triadic Componential Framework. All these various factors should be considered when designing pedagogical tasks, rather than simply relying on the manipulation of cognitive factors alone. Furthermore, measuring performance by way of CAF needs to be further developed so that more comprehensive pictures of learners' performance can be obtained.

REFERENCES

- Ahmadian, J. M., & Tavakoli, M. (2010). The effects of simultaneous use of careful online planning time and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15:1, pp. 35-59.
- Anton, M. (1999). The discourse of a learner centered classroom: sociocultural perspectives on teacher-learner interaction in the second language classroom. *The Modern Language Journal*, 83:3, pp. 303-318.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Beaven, B. (2005). Task-based approaches to teaching and learning languages. *Cornelson Verlag GmbH & Co. OHG. Autumn 05*, pp. 1-4.
- Bell, C. D. S. (2003). L2 speech rate in monologic and dialogic activities. *Linguagem & Ensino*, 6:2, pp. 55-79.
- Brumfit, C., & Johnson, K. (1979). *The communicative approach to language teaching*. London: Oxford University Press.
- Bygate, M. (1999). Task as the context for framing, re-framing and unframing of language. *System*, 27, pp. 33-48.
- Bygate, M. (2009). Effects of task repetition on the structure and control of oral language. In Van den Branden, K., Bygate, M., & Norries, J.M. (Eds.), *Task-based language teaching. A reader* (pp. 249-274). Amsterdam: John Benjamins Publishing Company.
- Bygate, M., Skehan, P., & Swain, M. (2001). *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Longman.
- Bygate, M., Cook, G., Iannou-Giorgiou, S., & Julian, P. (2003). Debate and discussion: Tasks are nothing new: They are just exercises with a name. In Pulverness, A. (Ed.), *IATEFL 2003 brighton conference selections*. Brighton, U.K.: IATEFL.
- Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Language learning & language teaching: Planning and task performance in a second language* (pp. 37-76). Amsterdam: John Benjamins Publishing Company.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1:1, pp. 1- 47.

- Candlin, C. N. (1987). Towards task-based language learning. In Candlin, C. N, & Murphy, D. (Eds.), *Language learning tasks* (pp. 5-22). Englewood Cliffs: N.J.: Prentice-Hall.
- Chuang, Y. Y. (2009). Foreign language speaking assessment: Taiwanese college English teachers' scoring performance in the holistic and analytic rating methods. *The ASIAN EFL Journal*, 11:1, pp.150-173.
- Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). New York: Routledge.
- Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33:2, pp. 185-209.
- Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, N.J.: Pearson Merrill Prentice Hall.
- Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative Research* (3rd ed.). Upper Saddle River, N.J.: Pearson Merrill Prentice Hall.
- Creswell, J. W. (2009). *Research design. Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Upper Saddle River, N.J.: Pearson Merrill Prentice Hall.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Crookes, G. (1986). *Task classification: A cross-disciplinary review* (Tech. Rep. No. 4). Honolulu: University of Hawaii at Manoa, Social Science Research Institute, Center for Second Language Classroom Research.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in second language acquisition*, 11, pp. 367-383.
- Crookes, G., & Long, M. H. (1987). Task-based second language teaching: A brief report. *Modern English Teacher*, 24:5, pp. 26-28.
- Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon, UK: Multilingual Matters.
- Davies, A. (2011). Does language testing need the native speaker? *Language Assessment Quarterly*, 8, pp. 291-308.

- De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, 13:1, pp. 1-24.
- Dornyei, Z. (2005). *The psychology of the language learner*. Mahwah NJ: Erlbaum.
- Duff, P. (1986). Another look at interlanguage talk: Talking task to task. In Day, R. (Ed.), *Talking to learn. Conversation in second language acquisition* (pp. 147-181). Rowley, M.A.: Newbury House.
- Elder, C., & Iwashita, N. (2005). Planning for test performance. In Ellis, R. (Ed.), *Language learning and language teaching: planning and task performance in a second language* (pp. 219-238). Amsterdam: John Benjamins Publishing Company.
- Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*. 4, pp. 193-220.
- Ellis, R. (2001). The differential effects of three types of task planning on the fluency, complexity and accuracy in L2 oral production. *Applied Linguistics*, 30:4, pp. 474-509.
- Ellis, R. (2003). *Task-based language teaching and learning*. Oxford: Oxford University Press.
- Ellis, R. & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, pp. 59-84.
- Ellis, R. (2005a). Instructed language learning and task-based teaching. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning* (pp. 713-728). New York; London: Routledge.
- Ellis, R. (2005b). Principles of instructed language learning. *System*, 33, pp. 209-224.
- Ellis, R. (2005c). Planning and task-based performance: Theory and research. In Ellis, R. (Ed.), *Language learning and language teaching: planning and task performance in a second language* (pp. 3-34). Amsterdam: John Benjamins Publishing Company.
- Ellis, R. (2009a). Task-based language teaching: Sorting out the misunderstanding. *International Journal of Applied Linguistics*. Vol. 19:3, pp. 221-246.
- Ellis, R. (2009b). The differential effect of three types of task planning on fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30:4, pp. 474-509.
- Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lesson. *Language Learning*, 51:2, pp.281-318.

- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Faerch, C., & Kasper, G. (1983). *Strategies in interlanguage communication*. London: Longman.
- Faerch, C., & Kasper, G. (1986). Cognitive dimensions of language transfer. In Kellerman, E., & Sharwood-Smith, M. (Eds.), *Cross-linguistic influence in second language acquisition* (pp. 20-60). Oxford: Pergamon Press.
- Fayer, J. M., & Krasinski, E. (1987). Native and non-native judgments of intelligibility and irritation. *Language Learning*, 37:3, pp. 313-326.
- Fernandez-Garcia, M. (2007). Tasks, negotiation, and L2 learning in a foreign language context. In Garcia-Mayo, M.P. (Ed.), *Investigating tasks in formal language learning* (pp.69-90). Clevedon: Multilingual Matters.
- Fillmore, C. (1979). On fluency. In Fillmore, C., Kempler, D., & Wang, W. (Eds.), *Individual differences in language ability and language behavior* (pp.85-102). New York, U.S.A: Academic Press.
- Finocchiaro, M., & Brumfit, C. (1983). *The functional-notional approach: From theory to practice*. New York: Oxford University Press.
- Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics*, 19:1, pp. 1-23.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, pp. 299-323.
- Foster, P., & Skehan, P. (2009). The influence of planning and task types on second language performance. In Van den Branden, K., Bygate, M., & Norries, J.M. (Eds.), *Task-based language teaching: a reader* (pp. 275-300). Amsterdam: John Benjamins Publishing Company.
- Foster, P., Tonkin, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21:3, pp. 354-375.
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59:4, pp. 866-896.
- Fujii, J., & Mackey, A. (2009). Interactional feedback in learner-learner interactions in a task-based EFL classroom. *International Journal of Applied Linguistics*, 47, pp. 267-301.
- Fulcher, G. (2003). *Testing second language speaking*. Pearson Education: Longman.

- Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Gass, S., Mackey, A., & Pica, T. (1998). The role of input and interactions in second language acquisition: Introduction to the special issue. *Modern Language Journal*, 82:3, pp.299-307.
- Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning*, 55:4, pp. 575-611.
- Gass, S., Mackey, A., & Ross-Feldman, L. (2011). Task-based interactions in classroom and laboratory settings. *Language Learning, Supplement*, 1, pp. 189 – 220.
- Gilabert, R. (2005). *Task complexity and L2 narrative oral production*. (Unpublished doctoral dissertation). Universitat de Barcelona.
- Gilabert, R. (2007a). The simultaneous manipulation of tasks complexity along planning time and (+/- Here-and-Now): Effects on L2 oral production. In Garcia Mayo, M. P. (Ed.), *Investigating tasks in formal language teaching: SLA* (pp.44-68). Clevedon, UK: Multilingual Matters.
- Gilabert, R. (2007b). Effects of manipulating task complexity on self-repairs during L2 oral production. *Review of Applied Linguistics in Language Teaching*, 45:3, pp. 215-240.
- Gilabert, R., Baron, J., & Llanes, A. (2009). Manipulating cognitive complexity across task types and its impact on learner's interaction during oral performance. *Review of Applied Linguistics in Language Teaching*, 4:3, pp. 367-396.
- Griffiths, R. (1991). Pausological research in an L2 context: a rationale and review of selected studies. *Applied Linguistics*, 12:4, pp. 345-364.
- Haden, B. L. (1991). Teacher and non-teacher perceptions of second language communication. *Language Learning*, 41, pp. 1-24.
- House, J. (1996). Developing pragmatic fluency in English as a foreign language. *Studies in Second Language Acquisition*, 18, pp. 225-255.
- Housen, A., & Kuiken, V. (2009). Complexity, Accuracy, and Fluency in second language acquisition. *Applied Linguistics*, 30:4, pp. 461-473.
- Ishikawa, T. (2007). The effect of manipulating task complexity along the (+/-here-and-now) dimension on L2 written narrative discourse. In Garcia Mayo, M. P. (Ed.), *Investigating tasks in formal language teaching: SLA* (pp.136-156). Clevedon: Multilingual Matters.

- Iwashita, N. (2001). The effect of learner proficiency on interactional moves and modified output in nonnative interaction in Japanese as a foreign language. *System*, 29, pp. 267-287.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23:1, pp. 83-103.
- Kasihani, K. (2010). Pengajaran bahasa inggris di sekolah dasar: Kebijakan, implementasi, kenyataan. *Orasi ilmiah pengukuhan guru besar*. Universitas Negeri Malang. *Pdf-Finder.com*. Google Scholar.
- Kim, H. J. (2005). *World Englishes and language testing: the influence of rater variability in the assessment process of English language oral proficiency*. (Unpublished Doctoral Dissertation,). The University of Iowa.
- Kim.Y. (2009). The effect of task complexity on learner-learner interaction. *System*, 37:2, pp. 254-268.
- Kitao, S. K., & Kitao, K. (1996). testing communicative competence. *The Internet TESL Journal*, 2:5. Retrieved from <http://iteslj.org/Articles/Kitao-testing.html>.
- Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning*, 49:2, pp.303-342.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, 62:2, pp. 439-472.
- Kuiken, F., & Vedder, I. (2007). Cognitive tasks complexity and linguistic performance in French L2 writing. In Garcia Mayo, M. P. (Ed.), *Investigating tasks in formal language teaching: SLA* (pp.117-135). Clevedon: Multilingual Matters
- Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17, pp. 48-60.
- Larsen-Freeman, D. (2000). *Techniques and principles in language teaching: teaching techniques in English as a second language* (2nd ed.). Oxford: Oxford University Press.
- Larsen-Freeman, D. (2009). Adjusting expectations: the study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30:4, pp.579-589.

- Lauder, A. (2008). The status and function of English in Indonesia: a review of key factors. *Makara, Sosial Humaniora*, 12:1, pp.9-20.
- Lennon, P. (1990). Investigating fluency in EFL a quantitative research. *Language Learning*, 40:3, pp. 387-417.
- Lennon, P. (2000). The lexical element in spoken second language fluency. In Reiggenbach, H. (Ed.). *Perspective on fluency* (pp. 25-42). Michigan: The University of Michigan Press.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Massachusetts Institute of Technology Press.
- Long, M. H. (1983). Native speaker/non-native speaker conversation in the second language classroom. In Clarke, M., & Handscombe, J. (Eds.), *Pacific perspectives on language and teaching* (pp. 207-225). Washington, DC: TESOL.
- Long, M. H. (1985). Input and second language acquisition theory. In Gass, S.M., & Madden, C.G. (Eds.). *Input in second language acquisition* (pp. 377-393). Cambridge MA: Newbury House.
- Long, M. H. (1991). Focus on Form: A design feature in language teaching methodology. In K. de Bot, R.B. Ginsberg, and C. Cramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins Publishing Company.
- Long, M. H. (1996). The Role of the linguistic environment in second language acquisition. In Ritchie, W.C., & Bathia, T.K. (Eds.), *Handbook of second language acquisition* (pp. 413-486). New York: Academic Press.
- Long, M.H. (2000). Focus on Form in task-based language teaching. In Lambert, R.D., & Shohamy, E. (Eds.), *Language policy and pedagogy: Essays in honour of Walton, A. R.* (pp.179-191). Amsterdam: John Benjamins Publishing Company.
- Long, M.H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26:1, pp. 27-56.
- Long, M.H., & Robinson, P. (1998). Focus on Form. Theory, research, and practice. In Doughty, C., & Williams, J. (Eds.), *Focus on Form in classroom second language acquisition* (pp.15-41). Cambridge: Cambridge University Press.
- Long, M.H., & Norris, J.M. (2009). Task-based teaching and assessment. In Van den Braden, K., Bygate, M., & Norris J.M. (Eds.), *Task-based language teaching. A reader* (pp. 135-142). Amsterdam: John Benjamins Publishing Company.
- Louma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

- Luciana. (2005). Negotiation of meaning in communicative tasks. *Indonesian Journal of English Language Teaching*, 1:1, pp. 45-56.
- Lyster R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19, pp. 37-66.
- Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, 30, pp. 459-477.
- Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: an exploration of NS-NNS and NNS-NNS adults and child dyads. *Language Learning*, 53:1, pp. 35-66.
- Mackey, A., & Silver, R.E. (2005). Interactional tasks and English L2 learning by immigrants in Singapore. *System*, 33, pp. 239-260.
- Mackey, A., Kanganas, A.P., & Oliver, R. (2007). Task familiarity and interactional feedback in child ESL classrooms. *TESOL QUARTERLY*, 41:2, pp. 285-311.
- MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk (3rd ed.). Mahwah, N.J.: Erlbaum.
- McDonough, (2004). Learner-learner interaction during pair and small group activities in a Thai EFL context. *System*, 32, pp. 207-224.
- McKay, G. (2004). Phonology: The organisation of speech sounds. In Conlan, C. (Ed.), *Teaching English language in Australia. Theoretical perspectives and practical issues* (pp. 69-84). Australia: API Network.
- McNamara, T. F. (1996). *Measuring second language performance*. Addison Wesley: Longman Limited.
- Mehnert, U. (1998). The Effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, pp. 52- 83.
- Michel, C. M., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL*, 45:3, pp. 241- 259.
- Mickan, P. (2004). Teaching methodologies. In Conlan, C. (Ed.). *Teaching English language in Australia. Theoretical perspectives and practical issues* (pp. 171-192). Australia: API Network.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded source book* (2nd ed.). Thousand Oaks, CA: SAGE.
- Na, Z. (2007). A study of high school students' English learning anxiety. *The Asian EFL Journal*, 9:3, pp. 22-34.

- Nabei, T., & Swain, M. (2010). Learner awareness of recasts in classroom: A case study of an adult EFL student's second language learning. *Language Awareness, 11:1*, pp. 43-63.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The Case of complexity. *Applied Linguistics, 30:4*, pp. 555-578.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge [England]: Cambridge University Press.
- Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press.
- Nunan, D., & Keobe, K. (1995). Task difficulty from learners' perspective: perceptions and reality. *Hong Kong papers in linguistics and language teaching, 18*, pp. 1-12.
- Ochs, E. (1979). Planned and unplanned discourse. In Givon, T. (Ed), *Syntax and semantics: discourse and syntax* (pp. 51-80). New York: Academic Press.
- Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition, 17*, pp. 459-481.
- Oliver, R. (1998). Negotiation of meaning in child interactions. *The Modern Language Journal, 82:3*, pp. 372-386.
- Oliver, R. (2000). Age differences in negotiation for meaning in child interactions. *Language Learning, 50:1*, pp. 119-151.
- Oliver, R. (2002). The patterns of negotiation for meaning in child interactions. *Modern Language Journal, 86:1*, pp. 97-111.
- Oliver, R., Grote, E., Rochecouste, J., & Exell, M. (2013). Needs analysis for task-based language teaching: a case study of indigenous vocational education and training students who speak EAL/EAD. *TESOL in Context, 22:2*, pp. 36-50.
- Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In Tashakkori, A., & Teddlie, C. (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-383). Thousand Oaks, CA: SAGE.
- Ortega, L. (1999). Planning and Focus on Form in L2 oral performance. *Studies in Second Language Acquisition, 21*, pp. 109-148.
- Oxford, R. L. (2006). Task-based language teaching and learning: An overview. *Asian EFL Journal: English Language Teaching and Research Articles, 8:3*, pp. 1-18.

- Pallant, J. (2007). *SPSS survival manual: a step by step guide to data analysis using SPSS for Windows (Third Ed.)*. Maidenhead, Berkshire: Open University Press McGraw Hill.
- Palloti, G. (2009). Defining, refining, and differentiating constructs. *Applied Linguistics*, 30:4, pp. 590-601.
- Philp, J., Oliver, R., & Mackey, A. (2006). The impact of planning time on children task-based interactions. *System* 34, pp. 547-565.
- Pica, T. (1984). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, 6, pp. 69-78.
- Pica, T., Young, R., & Doughty, C. (1987). Impact of interaction on comprehension. *TESOL Quarterly*, 2:4, pp.737-758.
- Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language research and interaction. In Crookes, G., & Gass, S.M. (Eds.), *Task-based learning in a second language* (pp.1-22). Clevedon: Multilingual Matters.
- Prabhu, N. (1987). *Second language pedagogy*. Oxford: Oxford University Press.
- Preston, C.C., & Colman, A.M. (2000). Optimal number of response in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, pp. 1-15. Elsevier.
- Rahim, P.R.M.A, Ahmad, N., & Rosly, N. (2004). Task-based approach as a teaching strategy in promoting the speaking skill. In Mukundan, J., Abidin, D. Z., & Hussin, A. A. (Eds.), *ELT Matters 1: Issues in English language learning and teaching* (pp. 173-180). Selangor [Malaysia]: University Putra Malaysia Press.
- Riggenbach, H.(1989). *Nonnative fluency in dialogue versus monologue speech: A microanalytic approach*. (Unpublished doctoral dissertation). University of California, Los Angeles.
- Richards, J.C., & Rodgers, T.S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge: Cambridge University Press.
- Robinson, P. (1995a). Task complexity and second language narrative discourse. *Language Learning*, 45, pp. 99-140.
- Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA. In Robinson, P. (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge: Cambridge University Press.

- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22:1, pp. 27-57.
- Robinson, P. (2003). The cognitive hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21:2, pp. 45-105.
- Robinson, P. (2005). Cognitive complexity and task sequencing: studies in a componential framework for second language task design. *Review of Applied Linguistics in Language Teaching*, 43, pp. 1-32.
- Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In Garcia Mayo, M.P. (Ed.), *Investigating tasks in formal language teaching: SLA* (pp.7-27). Multilingual Matters.
- Robinson, P. (2007b). Task complexity, theory of mind, and intentional reasoning: effects on L2 speech production of task difficulty. *Review of Applied Linguistics in Language Teaching*, 45:3, pp. 193-213.
- Robinson, P. (2011). Task-based language learning: a review of issues. *Language Learning. Supplement, 1*, pp. 1-36.
- Saldana, J., (2009). *The coding manual for qualitative researchers*. London: SAGE.
- Sajavara, K. (1987). Second language speech production: factors affecting fluency. In Dechert, H.W., & Raupach, M. (Eds.), *Psycholinguistics models of production* (pp. 45-65). Norwood, N.J.: Ablex.
- Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Palgrave, London: Macmillan.
- Saragih, A. (2009). Penguasaan bahasa Inggris lemah. *Harian Sumut Pos*. Retrieved 22 November, 2010, from <http://www.hariansumatraposts.com/2009/12/23408>.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, pp. 11-26.
- Schmidt, R. (2001). Attention. In Robinson, P. (Ed.), *Cognition and second language acquisition* (pp. 3-32) Cambridge: Cambridge University Press.
- Setyadi, B. (2009). Pembelajaran bahasa Inggris gagal. *Harian Lampung Post*. Retrieved 25 November, 2009, from [http:// www.lampungpost.com](http://www.lampungpost.com).
- Shehadeh, A., & Coombe, C. (2010). Introduction: from theory to practice in task-based learning. In Shehadeh, A., & Coombe, C. (Eds.). *Application of task-based language learning in TESOL* (pp. 1-7). Alexandria, VA: TESOL International Association.

- Skehan, P. (1996). A Framework for the implementation of task-based instruction. *Applied Linguistics*, 17, pp. 38-62.
- Skehan, P. (1998). *A Cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, pp. 1-14.
- Skehan, P. (2009). Modeling second language performance: integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, 30:4, pp. 510-523.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retelling. *Language Learning*, 49:1, pp. 93-120.
- Storch, N. (2002). Patterns of interactions in ESL pair work. *Language Learning* 52:1, pp. 119-158.
- Swan, M. (2005). Legislation by hypothesis: the case of task-based instruction. *Applied Linguistics*, 26:3, pp. 376 – 401.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: SAGE.
- Tavakoli, P. (2009). Investigating task difficulty: Learners' and teachers' perceptions. *International Journal of Applied Linguistics*, 19:1, pp. 1-25.
- Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative on learner output. *Language Learning*, 1, pp. 37-72.
- Tsai, C. I. (2003). Anxiety and beliefs about language learning: a study of Taiwanese college students learning English. (Unpublished Doctoral Dissertation). Texas A & M University-Kingsville, USA.
- Van den Branden, K. (2006). *Task-based language education. From theory to practice*. Cambridge: Cambridge University Press.
- Van den Branden, K., Bygate, M., & Norris, J. M. (2009). *Tasked-based language teaching: introducing the reader*. Amsterdam: John Benjamins Publishing Company.
- Winke, P., Gas, S., & Myford, C. (2012). Raters' L2 background as a source of bias in rating oral performance. *Language Testing*, 30:2, pp. 231-252.
- Wolfe-Quintero, K., Inagaki, S., and Kim, H. Y. (1998). *Second language development in writing: measures of fluency, accuracy, and complexity*. Honolulu: University of Hawaii.

- Yuan, F., & Ellis, R. (2003). The effect of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics*, 24, pp. 1-27.
- Yuan, F., & Ellis, R. (2005). The effect of careful within-task planning on oral production and written task performance. In Ellis, R. (Ed.), *Language learning & language teaching: planning and task performance in second language* (pp. 167-192). Amsterdam: John Benjamins Publishing Company.
- Yufrizal, H. (2001). *Negotiation of meaning in EFL dyads in Indonesia: Roles of gender, proficiency, and learning styles*. (Unpublished doctoral thesis). La Trobe University, Victoria.
- Zein, M. S. (2010). Educating primary school English teachers in Indonesia: primary study. Retrieved 9/12/2010. *TEFLIN Journal in Indonesia*. pp. 1-34.
- Zhang, Y., & Elder, C. (2010). Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing*, 28:1, pp. 31- 50.

APPENDICES

Appendix 1

The planned Simple Task (Task 1)

Both of you are planning to buy a mobile phone (Blackberry) as a gift for your friend's birthday. Please discuss in pairs the two different types of blackberries, comparing and arguing every detailed specification of the two types of Blackberries. Which would be the best mobile phone for you to buy.

Blackberry Onyx white



Specifications:

Price: Rp.4.450.000

Size (mm): length:109, width: 60, and thickness: 14.

Wight: 133g

Ringtones: plyponic, MP3

Memory: card slot 16 GB

Features: Messaging:SMS, MMS, email

Camera: 3.00 Mega pixel

Battery: standby up to 480 hours.

Talk time: up to 6 hours

Blackberry Bold



Specifications:

Price: Rp.3.750.000

Size: dimensions (mm): length:114, width:66, thickness:14

Wight: 122g

Ringtones: polyponic, MP3

Memory: card slot 128MB

Features: Messaging:SMS, MMS, email

Camera: 3.15 Mega pixel

Battery: standby up to 310 hours.

Talk time: up to 5 hours

Appendix 2

The Unplanned Simple Task (Task 2)

Both of you are trying to find a house for rent and to share. Please discuss the two types of houses by comparing and giving your opinion of every single detailed specification of the two houses. Please decide which would be the best house for you to rent according to your discussion.

Arjuna Type



Facilities: 2 small bedrooms including desk and chairs , 1 guest room, kitchen, 1 bathroom, small front and rear gardens

Price: R

p. 600.000 per month *including* electricity and water

Location: thirty minutes walking to university, close to bus station, and a shopping centre.

Bima Type



Facilities: 2 large bedrooms including mattress, 1 guest room, kitchen, 2 bathrooms, small front and large rear gardens.

Price: Rp. 750.000 per month' *excluding* electricity and water

Location: ten minutes walking to university, close to a sport center, and a restaurant

Appendix 3

The Planned Complex Task (Task 3)

Both of you are planning to buy a mobile phone (Blackberry) as a gift for your friend's birthday. Please discuss in pairs the six different types of Blackberries, comparing and arguing every detailed specification of the two types of Blackberries. Which would be the best mobile phone for you to buy.

<p>Blackberry 8830</p>  <p>Price: Rp.2750.000 Color :Blue Camera: 2.0 Mega Pixel Size (mm): length:1113, width: 75, and thickness: 20. Weight: 125 g Features: SMS & Email Ringtone: Polyphonic, No MP3 Memory: RAM:1MB Battery: standby:380 hours Talk Time: 5 hours</p>	<p>Blackberry 6230</p>  <p>Price: Rp.3000.000 Color :Red Camera: 2.15 Mega Pixel Size (mm): length:114, width: 66, and thickness: 14. Weight: 120 g Features: SMS, MMS, Email Ringtone: Polyphonic, MP3 Memory: RAM:1MB Battery: standby: 350 hours Talk Time: 4 hours</p>	<p>Blackberry Pearl 8100</p>  <p>Price: Rp.3500.000 Color : White Camera: 2.50 Mega Pixel Size (mm): length:107, width: 50, and thickness: 15. Weight: 90 g Features: SMS, MMS, Email Ringtone: Polyphonic, MP3 Memory: 1 MB Battery : standby: 360 Talk Time: 4.5 hours</p>
<p>Blackberry 8520</p>  <p>Price: Rp.4000.000 Color :Black Camera: 3.0 Mega Pixel Size (mm): length:110, width: 60, and thickness: 13. Weight: 115 g Features: SMS, MMS, Email Ringtone : Polyphonic, MP3, IM Battery: standby: 400 Talk Time : 5 hours</p>	<p>9650 Bold</p>  <p>Price: Rp.450.000 Color :Black Camera: 3.5 Mega Pixel Size(mm): length:114, width: 65, and thickness: 14. Weight: 110 g Features: SMS, MMS, Email, IM, Games Ringtone Polyphonic, MP3,WMA9 Battery: standby:450 hours Talk Time : 5 hours</p>	<p>Bb Curve 3G</p>  <p>Price: Rp.5000.000 Color :Black Camera: 4.00 Mega Pixel Size (mm): length:109, width: 60, and thickness: 14. Weight: 118 g Features: SMS, MMS, Email, IM, Games Ringtone: Polyphonic, MP3,WMA9 Battery: standby: 500 hours Talk Time : 5.5 hours</p>

Appendix 4

The Unplanned Complex Task (Task 4)

Both of you are trying to find a house for rent and to share. Please discuss the six types of houses by comparing and giving your opinion of every single detailed specification of the six houses. Please decide which would be the best house for you to rent according to your discussion.

<p>Type 22/70</p>  <p>Facilities: 1 large bedroom with a fan, 1 studying room including desks and chairs, 1 bathroom, kitchen, and no car port Location: close to the post office and university Price: Rp. 400.000/month <i>excluding</i> electricity and water</p>	<p>Type 30/78</p>  <p>Facilities: 2 small bedrooms, lounge including desks and chairs, 1 bathroom, kitchen, and no carport Location: close to bus station and local library Price: Rp. 600.000/month <i>including</i> electricity and water</p>	<p>Type 36/84</p>  <p>Facilities: 2 bed rooms with matrass, a lounge, 2 bathrooms, kitchen, and carport Location: close to bus station and shopping center Price: Rp 700.000/month <i>including</i> electricity and water</p>
<p>Superior Type 30/70</p>  <p>Facilities: 1 large bedroom with matrass, 1 studying room including desks and chairs, a lounge, a kitchen, 1 bathroom, and a carport Location: close to university and bus station Price: Rp. 500.00/month <i>excluding</i> electricity and water</p>	<p>Deluxe Type 40/91</p>  <p>Facilities: 2 small bedrooms with matrass, a lounge, a kitchen, a bath room, and a carport Location: close to sport center and bus station price: Rp. 700.000/month <i>excluding</i> electricity and water</p>	<p>Mansion Type 50/120</p>  <p>Facilities: 2 large bed rooms with matrass, a lounge with desks and chairs, a kitchen, 2 bathrooms, and carport Location: close to university and hospital Price: Rp 900.000/month <i>excluding</i> electricity and water</p>

Appendix 5

Interview Protocols for Indonesian Participants

1. What do you think the task you have performed? For example, whether the task was easy or difficult for you to do?
2. Why do you say this?
3. How do you feel while doing the task? For example whether you feel relaxed, frustrated, or satisfied, etc.
4. Why did you think made you feel like this?
5. How successfully (well) did you think you completed the task?
6. [Why?]
7. What is your opinion about the task? For example whether it is interesting or not interesting.
8. [Why?]
9. Do you think you would do the task like this again?
10. If so why, if not, why not?
11. Do you think that the tasks provide any learning opportunities?
12. [Why did you think this?]
13. Do you like doing the tasks interactively (having dialogue)?
14. [Why?]

Notes: as the participants are those majoring in English, questions were asked in English with Bahasa Indonesia (Indonesian Language) used for feedback loops and for clarification.

L1 Speakers' Rating Scales info sheet for Task 1

Speaker	Rating scales
Speaker 1 (A)	<div> 1 2 3 4 5 6 7 8 9 10 <div> _ _ _ _ _ _ _ _ _ </div> <div> very poor <div></div> very good </div> </div>
Speaker 2 (B)	<div> 1 2 3 4 5 6 7 8 9 10 <div> _ _ _ _ _ _ _ _ _ </div> <div> very poor <div></div> very good </div> </div>

Please comment as freely and as widely as you can on any aspects of this conversation.

Appendix 7

L1 Speakers' Rating Scales info sheet for Task 2

In this video recording you will hear two Indonesian speakers of English discussing the merits two different types of houses. Please rate each speaker in terms of how good you think his/her English is, where 10 indicates a very good speaker, and 1 indicates a very poor speaker.

Speaker	Rating scales
Speaker 1	<div style="text-align: center;">1 2 3 4 5 6 7 8 9 10</div> <div style="text-align: center;"> _ _ _ _ _ _ _ _ _ </div> <div style="text-align: center;">very poor very good</div>
Speaker 2	<div style="text-align: center;">1 2 3 4 5 6 7 8 9 10</div> <div style="text-align: center;"> _ _ _ _ _ _ _ _ _ </div> <div style="text-align: center;">very poor very good</div>

Please comment on the ratings you have given. For example, did you feel that one speaker was better than the other? If so, why do you think this speaker was better? (Please give specific examples of the speaker's usage which causes you think this. For example "asks question well".) Or do you think both speakers have approximately the same level of English? If so, what features of their talk caused you think this? (For example, you might think that Speaker 1 was better than Speaker 2 at, say, pronunciation; but Speaker 2 had a better vocabulary.)

Please comment as freely and as widely as you can on any aspects of this conversation.

Appendix 8

L1 Speakers' Rating Scales info sheet for Task 3

In this video recording you will hear two Indonesian speakers of English discussing the merits six different types of mobile phones (Blackberries). Please rate each speaker in terms of how good you think his/her English is, where 10 indicates a very good speaker, and 1 indicates a very poor speaker.

Speaker	Rating scales
Speaker 1	<div style="text-align: center;">1 2 3 4 5 6 7 8 9 10</div> <div style="text-align: center;"> _ _ _ _ _ _ _ _ _ _ </div> <div style="text-align: center;">very poor very good</div>
Speaker 2	<div style="text-align: center;">1 2 3 4 5 6 7 8 9 10</div> <div style="text-align: center;"> _ _ _ _ _ _ _ _ _ _ </div> <div style="text-align: center;">very poor very good</div>

Please comment on the ratings you have given. For example, did you feel that one speaker was better than the other? If so, why do you think this speaker was better? (Please give specific examples of the speaker's usage which causes you think this. For example "asks question well".) Or do you think both speakers have approximately the same level of English? If so, what features of their talk caused you think this? (For example, you might think that Speaker 1 was better than Speaker 2 at, say, pronunciation; but Speaker 2 had a better vocabulary.)

Please comment as freely and as widely as you can on any aspects of this conversation.

Appendix 9

L1 Speakers' Rating Scales info sheet for Task 4

In this video recording you will hear two Indonesian speakers of English discussing the merits six different types of houses. Please rate each speaker in terms of how good you think his/her English is, where 10 indicates a very good speaker, and 1 indicates a very poor speaker.

Speaker	Rating scales
Speaker 1	<div style="text-align: center;">1 2 3 4 5 6 7 8 9 10</div> <div style="text-align: center;"> _ _ _ _ _ _ _ _ _ _ </div> <div style="text-align: center;">very poor very good</div>
Speaker 2	<div style="text-align: center;">1 2 3 4 5 6 7 8 9 10</div> <div style="text-align: center;"> _ _ _ _ _ _ _ _ _ _ </div> <div style="text-align: center;">very poor very good</div>

Please comment on the ratings you have given. For example, did you feel that one speaker was better than the other? If so, why do you think this speaker was better? (Please give specific examples of the speaker's usage which causes you think this. For example "asks question well".) Or do you think both speakers have approximately the same level of English? If so, what features of their talk caused you think this? (For example, you might think that Speaker 1 was better than Speaker 2 at, say, pronunciation; but Speaker 2 had a better vocabulary.)

Please comment as freely and as widely as you can on any aspects of this conversation.

Appendix 10

Coding and in-Vivo Responses of the Participants' Perceptions of Task Difficulty

Question	Elicited Category	Code	In-Vivo Descriptors
1	Difficulty	1 +	easy, quite easy, not really difficult, not very difficult, not too difficult
		1 –	difficult, more difficult, bit difficult, hard
2	Stress	2 +	Relaxed, more relaxed, enjoyable
		2 –	not relaxed, not well, frustrated, little frustrated, confused, little confused, difficult, bit nervous, not satisfied, middle
3	Confidence	3 +	successful, rather successful, almost successful, better, well, quite well, very well, good enough, well enough
		3 –	not successful, unsuccessful, less successful, rather successful, not quite successful, not prepared, not well, not really good, not really well
4	Interest	4 +	interesting, quite interesting, good
		4 –	not interesting, bit interesting, just so-so
5	Motivation	5 +	yes, good, maybe, sometimes
		5 –	no, enough
6	Learning Opportunities	6 +	yes
		6 –	no, difficult, confused
7	Interactive Task	7 +	yes
		7 –	no

Appendix 11

Criteria of Acceptance for L1 Ratings Scales

No	Included ratings	Example	Task 1		Task 2		Task 3		Task 4	
			F	%	F	%	F	%	F	%
1	Three ratings are the same	5 5 5								
2	Two ratings are the same and the third within one or two-Likert Scale	5 5 4								
3	Three ratings within one-Likert Scale	3 4 5								
4	The two ratings are within one-Likert scale and the third rating is within two-Likert scale.	4 5 7								
	Total									
	Excluded ratings									
1	Two ratings are the same and the third within three-Likert scale	2 5 5								
2	Two ratings within one-Likert Scale and the third within three-Likert Scale	2 5 6								
3	Two ratings within two-Likert Scale and the third within two or three-Likert Scale	2 5 7								

