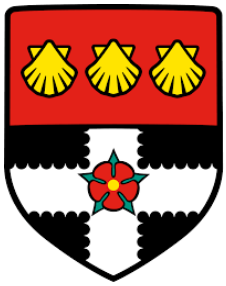

Improving Speaking Performance and L2 Motivation through Task-Based Language Teaching on Malaysian Undergraduate Students

Muhammad Yasir Yahya

Thesis submitted for the degree of

Doctor of Philosophy

Institute of Education

University of Reading

October 2019

DECLARATION

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Muhammad Yasir Yahya

31/10/2019

ABSTRACT

The present study investigated the effect of two different language teaching methods (TBLT and TSLT) on students' speaking performance and motivation towards learning and speaking English. The study employed a mixed-method, quasi-experimental, pre-/post-test design. The participants were 59 English as a second language (ESL) undergraduate students (aged 20 to 23), who were assigned to two intact groups namely the EIST (explicit instruction + speaking tasks, i.e. TSLT) group and the ST (speaking tasks only, i.e. TBLT) group. The instruments utilised within this study involved both quantitative and qualitative methods. Students' speaking performance was assessed through a speaking test whereas their motivation towards learning and speaking English was elicited through a questionnaire and interview. These instruments were implemented at the pre- and post-test data collection points before and after the speaking intervention which lasted for 8 weeks. For the data analyses, eleven measures of speaking and five sub-scales for motivation were employed to assess students' speaking performance (complexity, accuracy and fluency) and L2 motivation (attitude, integrativeness, instrumentality, linguistic self-confidence and learning situation), respectively.

The findings of the study indicated that students in both groups had an approximately equal level of speaking performance and motivation at the pre-test. However, at post-test, the results showed that the two teaching methods had different impacts on students' speaking performance and motivation. The TSLT method had a significant effect on the EIST group's vocabulary (lexical complexity) outcomes whereby the language produced was more diverse than the ST group. However, experiencing the TBLT method, the ST group managed to produce more complex and sophisticated language with the presence of abstract words, even though overall levels of lexical complexity and productivity were lower. In addition, the ST group had become more fluent over time with a significant decrease in pauses, repetition and reformulation

compared to the EIST group. Nevertheless, neither group demonstrated a statistically significant improvement across the measures of syntactic complexity and accuracy. As for motivation, the questionnaire data indicated that the two language teaching methods had no effect on students' attitudes, integrativeness and instrumentality. However, TSLT had a positive impact on the EIST group's linguistic self-confidence and learning situation motives, whereas TBLT had a positive effect on the ST group's linguistic self-confidence only. Nevertheless, within the interview, analysed qualitatively, students in both groups showed positive views across all sub-scales of motivation.

The study provides evidence that the two different language teaching methods affected students' speaking performance and L2 motivation in different ways. These results imply that the TSLT approach had a positive impact on students' overall vocabulary outcomes and L2 retention and the TBLT approach was more effective in promoting complex and fluent language. Furthermore, the study highlights the importance of the teacher's role and the TSLT approach in encouraging students' overall L2 motivation.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, Professor Suzanne Graham and Dr. Rowena Kasprowicz for their invaluable guidance, patient and encouragement. Likewise, I would also like to express my sincere gratitude to my previous supervisors, Dr. Louise Courtney and Professor Xiao Lan Curdt-Christiansen. I could not have gone through the struggles of completing my thesis without their endless support. It has been a real pleasure to work with them and I am deeply indebted to all of them.

I warmly thank Professor Jeanine Treffers-Daller for her assistance and valuable advice for the vocabulary analyses and the lemmatisation section. My sincerest thanks also go to Dr. Anthony Zhang for his insightful feedback and statistical advice.

My profound appreciation goes to my former colleagues Associate Professor Dr. Raha Mohd Radzuan and Zarina Mohd Ali from the Centre for Modern Languages and Human Science, Universiti Malaysia Pahang, who had given me the opportunity and assistance to conduct my study. My appreciation is also extended to Abdul Jalil, Norhayati, Sheikha and Faizahani for their support, time and efforts they had generously invested for this study. I would like to express special thanks to my PhD colleagues especially from the Institute of Education (L33).

The completion of my PhD journey would not have been possible without the financial support received from MARA and Faculty of Modern Languages and Communication, Universiti Putra Malaysia which is gratefully acknowledged.

Finally, my deepest gratitude goes to my wonderful parents, Yahya Che Lah and Nor Hashimah Hashim for their unconditional love and continuous *doa*, And last but not least to my lovely brothers, Yusri, Yazid, Yaqin and Hakiim and sisters-in-law, Shakila and Soraya Lin, thank you for the love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

1.1 Context of the Study

The context of the present study is Malaysia, which is situated in Southeast Asia. Malaysia is divided into two major geographical areas, West Malaysia (Peninsular) and East Malaysia (Borneo). The population of Malaysia as reported today is 32.4 million which comprises three main ethnic groups, namely *Bumiputera* (Malays and indigenous ethnic groups) forming the majority of the population which is approximately 69.1%, followed by Chinese, 23%; and Indian, 6.9% (Department of Statistics Malaysia, 2018). A total of 140 languages are spoken, hence Malaysians are either bilingual, trilingual or even multilingual (Darmi & Albion, 2013; David, Milde, & Coluzzi, 2009). Reflecting Malaysia as a multilingual nation, *Bahasa Melayu*/*Malaysia* (BM) or the Malay language has been made the national language that is used to unify the multicultural Malaysians especially through education (Muslim, 2013). The role of BM has somehow affected other languages in Malaysia especially the English language. The following sections give an overview of Malaysian education and how the English language became significant in the Malaysian education system.

1.1.1 The Overview of Educational System in Malaysia

The formal national education system in Malaysia is overseen by the Ministry of Education (MoE) and runs from the ages of 4 to 19 in which there are two or three years of pre-school (ages 4 to 6), six years of primary school (ages 6 to 12), five years of secondary school (ages 13 to 17) and one or two years of post-secondary (ages 18-19).

Primary education consists of two levels: Level One for pupils aged 6 to 9 (Year 1 to 3) and Level Two for pupils aged 10 to 12 (Year 4 to 6). For primary education, there are two types

of primary school: national-type schools where BM is used and taught, and vernacular schools in which the Mandarin or Tamil language is used as the medium of instruction. Historically, before Malaysia achieved independence from the British in 1957, the education system was divided according to three common ethnicities: Malay, Chinese and Indian. Each ethnic group established their own school systems to accommodate students whose first language was similar to the language of instruction (Abdul Rahman, 2014). Even after independence, Malaysia continues to practise linguistic segregation, evidenced by the existence of vernacular primary school until today (Darus, 2010). BM remains as the medium of instruction used in national schools while in Chinese schools as Mandarin language and in Indian schools as Tamil language. As for the English language, it is considered as a second language (ESL), taught in school. At the primary level, the curriculum encompasses seven core subjects (BM, English, Mathematics, Science, History, Civics and Citizenship Education, and Islamic or Moral Education) and four compulsory subjects (Physical and Health Education, Art and Design, Music, and Computing and Technology). Before progressing to secondary education, Year 6 pupils are required to sit the Primary School Achievement Test (*Ujian Pencapaian Sekolah Rendah*, UPSR) which only tests the Malay language comprehension, the writing skills of the Malay language, comprehension of English language, the writing skills of English language, Mathematics and Science. In addition to these six papers, Mandarin comprehension and writing are compulsory in Chinese schools and Tamil comprehension and writing in Indian schools.

Secondary education consists of Lower Secondary for students aged 13 to 15 (Form 1 to 3) and Upper Secondary for students aged 16 to 17 (Form 4 and 5). At the end of Form 3, students are required to take Form Three Assessment (*Pentaksiran Tingkatan 3*, PT3) which tests eight core subjects: BM, English, History, Geography, Science, Mathematics, Islamic Studies and Living Skills. The PT3 is used to place the students either into academic stream (science or arts), technical and vocational stream, or religious stream in for their upper secondary education.

Before graduating from secondary school, students have to sit for the Malaysian Certificate of Education (*Sijil Pelajaran Malaysia*, SPM). The type and number of subjects taken for SPM vary on the stream the students are in, but the core subjects for all are the same, which are BM, English, Islamic/Moral Studies, History, Mathematics and Science. In order to obtain the SPM certificate and graduate from secondary school, the MoE has made BM and History as compulsory subjects to pass.

In post-secondary education, students may opt to continue studying in Form 6 for one a half year or matriculation for a year. For Form 6, students are required to take the Malaysian Higher School Certificate (*Sijil Tinggi Persekolahan Malaysia*, STPM), which is administered by the Malaysian Examination Council (MEC). STPM is internationally recognised and is equivalent to the British A-Level examination. In addition to this post-secondary education, all students are obliged to take the Malaysian University English Test (MUET), a platform to test English proficiency amongst pre-university students prior to entering university (Michael, 2012).

In tertiary education, all universities are under the jurisdiction of the MoE but each university has been granted autonomy that allows the institution to run independently. The undergraduate programmes are offered at public or private universities in Malaysia. These programmes usually take up three to five years depending on the disciplines. Majority of the courses are conducted in English and students are required to enrol in English for Specific Purposes (ESP) courses (e.g., English for Academic Purposes, Technical English, English for Business Communication etc.) which they have to pass (Pillai, Khan, Ibrahim, & Raphael, 2012). However, different universities offer different ESP courses for their own language outcomes, to cater students' needs and meet the demands of the global market.

1.1.2 The Language Policy Changes in Malaysia

Language policy is seen to be the main instrument in nation building and national identity (Tollefson & Tsui, 2003). The teaching of the English language in the Malaysia education system has undergone drastic changes due to the changes in language policy after Malaysia gained its independence (Too, 2017). These changes were influenced by several factors including cultural, political and governmental control (Pennycook, 2017; Vethamani, 2007).

English was considered the first language in Malaysia during British colonisation. After independence, BM was stipulated to be the national language of Malaysia. The implementation of BM was to 'Malaysianising' education as a national agenda in shaping Malaysian identity and racial integration (Malakolunthu & Rengasamy, 2012). Furthermore, these changes were initiatives to establish an education system with a Malaysian-oriented curriculum to unite the nation (Abdul Rahman, 2014). Therefore, all government institutions are required to use BM as the official language of communication while retaining the vernacular Chinese and Indian schools (Darmi & Albion, 2013). The move from using English to BM as the medium of instruction was done gradually and initiated by the art subjects that later completed the move in all subjects in 1982 (Pandian, 2002). This was followed by the conversion in tertiary education which started in 1983. However, due to the lack of resources in BM, several disciplines such as medicine, engineering and law continued using English in most aspects of teaching and learning. (Pandian, 2002). The old policy which only benefited the people in the urban areas and elite classes (Selvaraj, 2010) was replaced with the National Education policy in 1970 to create a new national identity, thus stressing BM as only the first language as a way to unify the Malaysians (Pandian, 2002). The role of English language was reduced and was only taught as a subject in schools which resulted in limited exposure to English (Kayad, 2015) and decline in the standard of English among students (Tharmalingam, 2012).

To move towards globalisation, being competent in English is seen as a commodity and Malaysians must be competently proficient in it to meet challenges internationally (Tharmalingam, 2012). Therefore, in early 2003, the teaching of Mathematics and Science in English (*Pengajaran dan Pembelajaran Sains dan Matematik dalam Bahasa Inggeris,* PPSMI) policy was introduced. The reform of the education system was made whereby English was implemented as the medium of instruction for the teaching of mathematics and science subjects in all government schools in Malaysia. The decision to teach these subjects in English triggered unease from many sectors of the public, with many felt that reintroducing English as a language of instruction had been hastily taken without proper research and planning (Tharmalingam, 2012). However, MoE continued to implement the new curriculum to meet the demands for progress in science and technology (Chan & Tan, 2006), to encourage high English language proficiency among students and for future workplace requirements (Mustapha, 2008). Moreover, it is believed that the English language was critical, and still is, for the economic and technological development of the country (Tharmalingam, 2012).

Nevertheless, after nearly a decade, teaching Mathematics and Science in English was discontinued. The policy was abolished in 2011 since it was not working well and received backlashes from the public. A few of the reasons were that Mathematics and Science teachers were educated in the Malay medium where English was taught as a subject, thus they lacked the required level of competency in English to teach Mathematics and Science which negatively affecting the learning of these two subjects (Michael, 2012). In addition, the gap between the students in the urban and the rural areas in Mathematics and Science subjects achievements were large and very often, students in the rural areas were seen to be left behind (Yang & Ahmad Ishak, 2012).

In 2012, a bilingual policy called Upholding the Malay Language and Strengthening the English language (*Memartabatkan Bahasa Malaysia Memperkasakan Bahasa Inggeris,*

MBMMBI) was introduced (MoE, 2014). The policy's main aim is to ensure that each child masters both BM and English well and fluently (Rashid, Abdul Rahman, & Yunus, 2017). In addition, it was in line with the new government's aspiration which is to make BM not only the medium of unity but also the main language of communication and science, alongside English, thus enabling students to compete and explore new knowledge worldwide (Thirusanku & Md Yunus, 2012). English is seen to have a vital role for educational, political, economic growth, national development and unity (Too, 2017), hence the MoE launched the National Education Blueprint 2013-2025 for preschool to post-secondary education and the National Education Blueprint 2015-2025 for higher education in 2013 and 2015, respectively. It was introduced in the context of raising international education standards, preparing students for the needs of the 21st. century (MoE, 2013, 2015b). In addition, both blueprints highlighted the importance of being "proficient in BM and English and encouraged to learn one additional global language" (MoE, 2013, p. 112; 2015b, p. 15). Therefore, by placing BM and English at the same level indicates that both languages are given the same recognition in the role of developing knowledge and communication skills.

1.1.3 English Language Teaching (ELT) in Malaysia

Following the implementation of the National Education policy in 1970, the English syllabus in schools was divided into three phases with each phase employing three different teaching approaches. These phases were 1) the structural-situational approach which was employed in the primary school syllabus; 2) the contextual-based teaching which was adopted in the lower secondary syllabus; and 3) the communicative syllabus in the upper secondary (Pandian, 2002). The reason behind this disparity was because of the different ad hoc committees that were in charge of the development of the language curriculum during that time (Pandian, 2002). The structural syllabus mainly focused on the learning of grammar deductively while neglecting the communicative aspects of the English language (Mansor & Hasan, 2015).

In 1980, the Malaysian education system was revamped in order to produce a "united, discipline and well-trained workforce" (Pandian, 2002, p. 39). The old English language curriculum changed from form-focused syllabuses to a communicative task-oriented situational approach and skills-based syllabus (Hall, 2015; Pandian, 2002) or better known today as task-based language teaching (TBLT). Malaysia was one of the first countries to adopt TBLT for instruction (Nunan, 2003) and the new syllabus introduced many features of TBLT which supported communicative and meaningful tasks especially through the implementation of the New Primary Schools Curriculum (*Kurikulum Baru Sekolah Rendah*, KBSR) in 1983 and the Integrated Secondary Schools Curriculum (*Kurikulum Bersepadu Sekolah Menengah*, KBSM) in 1989 (Hall, 2015; Pandian, 2002). The KBSR and KBSM ELT syllabuses incorporate the communicative approach in which activities that simulate real-life contexts and lessons were conducted by integrating the four skills (listening, speaking, reading and writing) (Pandian, 2002).

While Malaysia has embraced TBLT as an approach to teaching English, there are still uncertainties regarding its suitability and effectiveness within primary and secondary education since it is deemed not been employed effectively in the ESL classroom even with the 200-minute per week lessons (Mustafa, 2012). Until today, teachers still considered using the traditional approach of teaching, also known as the presentation-practice-production (PPP) (Choo & Too, 2012; Mustafa, 2012; Pandian, 2002). The teaching of English in school concentrates more on vocabulary skills, grammar, reading comprehension and writing as these components are tested in the UPSR and SPM examination, hence task selection such as cloze passages and multiple-choice questions are driven by the goal of obtaining high scores in examinations (Hall, 2015; Pandian, 2002). The examination-oriented culture is "deeply rooted in the socio-cultural history in Asia" (Butler, 2011, p. 46) and thus ESL classrooms become more teacher-centred and little attention is given to listening and speaking skills (Hall, 2015;

Pandian, 2002). Oral tests in school are still used to assess students' speaking ability, but teachers do not emphasize speaking as it does not contribute towards students' national examination results. Furthermore, oral tests are conducted simply for the sake of documentation purposes (Mat Hassan & Talib, 2013). Therefore, the intense attention on examinations had overshadowed the communicative approach which has been the core of the development of the KBSR and KBSM curriculum.

In order to keep up with globalisation, the National Education Blueprint 2013-2025 was announced and the old curriculum was replaced with the Primary School Standards-Based Curriculum (*Kurrikulum Standard Sekolah Rendah,* KSSR) and the Secondary School Standards-Based Curriculum (*Kurrikulum Standard Sekolah Menengah,* KSSM). The new curriculum highlights teaching and learning in the 21st century that incorporates higher order thinking skills (e.g. creation, evaluation, analysis and application) (MoE, 2015a, 2018). The new language curriculum requires teachers to apply and design classroom activities based on several strategies and approaches which include inquiry-based learning, project-based learning, student-centredness and collaborative learning rather than using the traditional approach or rote-learning (Sulaiman, Mohd Ayub, & Sulaiman, 2015). Unlike the previous curriculum which weighted more on summative assessment, the new language curriculum employs both formative and summative assessment which help to measure students' performance continuously and comprehensively throughout the years of schooling (MoE, 2015a, 2018). Therefore, this gives a comprehensive assessment of students' overall language proficiency.

In post-secondary education, students are required to take Malaysian University English Test (MUET). Generally, MUET is a compulsory English language proficiency test that is used as admission into local universities, which is administered by the Malaysian Examinations Council (MEC). The main objective of MUET is to measure students' English language proficiency upon entry into tertiary education (MEC, 2015) and to provide diagnostic

information for students' placement in English language proficiency courses (Rethinasamy & Chuah, 2011). All the four language skills are tested in the MUET examination, but they are not measured equally. Reading is considered the most essential with the highest weightage (45%), followed by writing (25%), and listening and speaking are both weighed the least (15%) (Kaur & Nordin, 2006; Rethinasamy & Chuah, 2011). The MUET scores are reported in six-band scales, 1 being the lowest whilst 6 being the best. These scores demonstrate the student's overall English proficiency including communicative ability, comprehension and task performance.

As for the tertiary level, there is no explicit English language policy that specifically defines how the English language should be taught in the classroom (Too, 2017). Therefore, each university offers different English language courses which adopt different teaching methods to meet the students' needs and prepare them for the global market. However, ELT at the university level is heavily influenced by the National Education Philosophy (see MoE, 2015b, p. 79) and the development of soft skills which include "integrated thinking skills and knowledge culture, language proficiency (BM, English and a global language), national identity, national unity, leadership, problem-solving skills, ethics and spirituality, and entrepreneurial mind set and readiness" (MoE, 2015b, p. 82). The emphasis on the English language varies in each university in Malaysia and can be seen in the different range of the English language courses offered. Among the courses are English language proficiency courses, English for academic purposes (EAP) courses and English for occupational purposes (EOP) courses (Too, 2017). Students who have achieved band 3 and below in their MUET may need to enrol in general English courses, whereas those who have achieved band 4 and above may enrol in elective English courses (Darmi & Albion, 2013). General English courses are mainly remedial English classes which focus on assisting students who are poor in English. For students who have successfully passed the minimum language requirement, they are

required to enrol in EAP and EOP courses related to their respective undergraduate programmes. The EAP courses are aimed at preparing students to familiarise themselves with academic writing (e.g. dissertation, research project etc.), whereas EOP courses are offered to students in specific disciplines and they are designed to prepare them for professional and workplace settings (e.g. business English, technical English for engineering etc.).

1.2 Problem Statement

Due to the changes in language policy and the education system being reformed, language teaching approach has been divided. Some teachers may avoid using the traditional methods and adopt new teaching approach to meet the 21st –century teaching and learning (e.g. critical thinking, problem-solving, digital literacy etc.), while others might still include PPP in their lessons as it is considered necessary for language acquisition.

Task-based language teaching or TBLT has become a widely used language teaching approach explored in the field of second language acquisition research (Ellis, 2003, 2018; Long, 2015; Skehan, 1996, 2018). TBLT is seen to be the type of approach that primarily focuses on tasks as the main tool for language teaching and promoting communicative, meaningful practice in the L2 (Ellis, 2009c; Long, 2015). Nunan (1993) described a task as "a piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is principally focused on meaning rather than form" (p. 59). In other words, TBLT emphasises meanings in communicative tasks in which students are required to use their own linguistic skills to accomplish the tasks (Ji & Pham, 2018).

In the context of Malaysia, the implementation of TBLT is limited and constrained by the "formal top-down traditional methods" (Nunan, 2003, p. 602), prompting teachers to incorporate linguistic features such as lexical forms and grammatical structures into oral and written tasks as the main objective of teaching and learning (Choo & Too, 2012; Hall, 2015; Ji

& Pham, 2018; Mustafa, 2012; Too, 2017). In the context of Malaysia, although the foundation of the education curriculum is based on task-oriented approach, explicit grammar and vocabulary instruction still play a major role in ELT, especially in the examination context where emphasis is given on reading and writing tasks (Butler, 2011).

The rationale of using explicit language teaching is that language acquisition involves students noticing specific language features (Schmidt, 1990, 2010), hence boosting students' attention to use the target language structure more saliently (Robinson, 2007). By employing explicit grammar and vocabulary instruction in communicative tasks, teachers could design tasks that could integrate and balance both the linguistic features and communication (Ortega, 2007). Therefore, Ellis (2003) proposed task-supported language teaching (TSLT) in which communicative tasks are used alongside explicit language instruction. Several studies have shown that explicit instruction is effective in promoting language acquisition and producing grammatically accurate use of the target language (Norris & Ortega, 2000; Spada & Tomita, 2010). However, some advocates of TBLT disagreed with employing explicit instruction (e.g. in pre-task stage), as it takes an approach similar to the traditional methods of ELT (Long, 2016; Skehan, 1996). It is believed that language instruction should be implicit and allow students to use their own linguistic resources (Long, 2016). Moreover, students may approach tasks as language exercise rather than as a meaningful communicative task if explicit instruction was adopted (Van de Guchte, Rijlaarsdam, Braaksma, & Bimmel, 2017). Therefore, the primary focus on meaning instantly becomes focus on forms (Ellis, 2016) and impedes the use of authentic language.

Task-based approaches are also believed to benefit and foster students' L2 motivation (Dörnyei, 2002; Ellis, 2003) in which students have the opportunity to have social interaction among their peers (Zúñiga, 2016) and are motivated in completing the given tasks (Brophy, 2004). Therefore, the high interest in language instruction and assessment have led second

language teachers and researchers to focus on the psychological traits of language learners (Neff, 2007). Theoretical constructs of motivation play an important role in SLA research and are believed to contribute to the success of second language learning.

Despite the efforts in the reformed of the Malaysian language curriculum to put value on communicative tasks, the decline of English language proficiency especially in speaking among students has been the focus of serious attention for the past few years (Che Musa, Koo, & Azman, 2012). The exam-oriented system that emphasises on grammar rules and rote-learning had made students feel demotivated to speak effectively and to be dependent on their teachers (Mansor & Hasan, 2015). This causes students to be ill-equipped to meet the communicative expectations for speaking determined by the university. Language instruction at the university level focuses on task completion rather than the development of the speaking skills and hence it gives impact on students speaking motivation. In a context where BM is predominantly used, students find speaking more difficult and problematic than any other language skills (Tunku Mohtar, Swaran Singh, Abdullah, & Mat, 2015). The problem is exacerbated when English is only heard and practised in the classroom and does not go beyond the classroom setting (Zhang, 2009). Therefore, the lack of exposure to English affects students' motivation to speak unless they are being prompted by the teacher or the teaching instruction and the exposure to communicative tasks. An important tenet of research into second language motivation is that different approaches and tasks arouse different motivational influences and thus affect the learning processes and the language outcome (Julkunen, 2001).

To sum up, there is a challenge and growing awareness among educators in Malaysia regarding the need to produce highly motivated English proficient students. The extent to which TBLT or TSLT facilitate this is worthy of exploration. Furthermore, investigating students' speaking performance also entails the study vocabulary, pronunciation, morphological and lexical, disfluencies and sentence boundaries (Muhamad et al., 2013). Therefore, the present study adds

to the existing body of literature by examining students' English speaking performance particularly in respect to its complexity, accuracy and fluency. Moreover, the connection between students' speaking performance and their motivation cannot simply be ignored as motivation plays a major role in students' success in speaking. Therefore, the aims and research questions of the present study have been designed to fill this gap.

1.3 Research Aims and Questions

In conjunction with investigating the issues highlighted in the preceding section, the aim of the present study is to investigate the effects of different forms of task-based language teaching (task-based and task-supported) on speaking performance in respect of complexity, accuracy and fluency, and L2 motivation towards learning and speaking the English language. To achieve the main aims, the following research questions of the study were developed.

1. What are the effects of the implementation of two different forms of language teaching (TBLT and TSLT) on:
    a. the development of students' English speaking performance over time?
    b. students' motivation towards the learning and speaking English over time?

1.4 Thesis Outline

This thesis has been organised into seven chapters. Chapter 1 had been presented in the context of the study and the significance of undertaking it.

Chapter Two reviews the literature pertaining to task-based language teaching (TBLT), the overview of the psycholinguistic model of speech production and the dimensions of the speaking (complexity, accuracy and fluency). In addition, this chapter discusses the second language motivation theories which include Gardners' Socio-Educational Model, Dörnyei's

Framework of L2 Motivation and Bandura's Self-Efficacy Theory, and also willingness to communicate.

Chapter Three details the research design and the methodological approaches employed in the study to answer the research questions of the study. To enrich the data from different perspectives, mixed method approaches were used for data collection and these instruments, including the intervention, were piloted before the actual study was conducted. The ethical considerations are also raised in this chapter.

Chapter Four presents the analyses and the findings of the effects of two different forms of language teaching (TBLT and TSLT) on students' speaking performance over time (pre- to post-test). The speaking data were transcribed and coded using CLAN software and analysed quantitatively using SPSS software. The speaking data were also analysed qualitatively, giving a more fine-grained analysis of students' speaking outcomes in terms of complexity, accuracy and fluency.

Chapter Five presents the analyses and findings of the effects of TSLT (EIST group) and TBLT (ST group) approach on students' motivation over time (pre- to post-test) obtained from students' questionnaires and interviews. The questionnaire was analysed using SPSS software and the results were presented quantitatively. Through thematic analysis, the interview data were sorted into five main themes, namely attitude, integrativeness, instrumentality, linguistic self-confidence and learning situation.

Chapter Six contains a discussion of the findings of the study in relation to the two research questions which the present study sought to answer, and in relation to existing research.

Chapter Seven concludes the overall study and main findings and reviews the limitations of the study. In conclusion, the study's implications for speaking and motivation development, the

contributions of the study to second language learning and recommendations for future research are also considered.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter reviews the literature relating to the present study. The study's primary aim is to explore the effects of two forms of language teaching (task-based and task-supported) on students' speaking performance and their second language motivation particularly in relation to speaking. Studies investigating the nature of those variables and how they impact language teaching and learning outcomes, as well as areas of debate concerning these variables, are also discussed in this chapter.

2.2 Task-Based Language Teaching (TBLT)

TBLT has emerged to be a subject of research interest for over three decades and it is still considered as one of the predominant teaching methods in the field of second language learning and teaching (Ur, 2013). Some researchers consider TBLT to be "the latest realization of Communicative Language Teaching (CLT)" (Nunan, 2003, p. 606), while others consider TBLT to be a post-method pedagogy not connected with any specific method (Kumaravadivelu, 2006). TBLT can generally be described as language pedagogy that focuses on communicative and collaborative tasks as the core for planning and delivery of instruction (Bygate, 2016; Long, 2016). It is based on the assumption that language is a means of communication and is effectively learnt through the exposure and negotiation that exist during the performance of communicative tasks (Ur, 2013). In TBLT, students are given authentic tasks which require them to focus mainly on exchanging meaning and using language for real-world purposes (Ellis, 2017). Therefore, the premise is that when students are task-focused and given the opportunity to practise authentic language in the classroom, they learn the target language better than when the focus of teaching is primarily (or exclusively) on language form.

2.2.1 Defining Task

Various definitions of task have been proposed under TBLT. For example, Skehan (1998) defined a task as an activity in which "meaning is primary; there is some communication problem to solve; there is some sort of relationship to comparable real-world activities; task completion has some priority; and the assessment of the task is in terms of outcome" (p.95). Similarly to Skehan, Long (2016) described task as "real-world communicative uses to which learners will put the L2 beyond the classroom - the things they will do in and through the L2 - and the task syllabus stands alone, not as one strand in a hybrid of some kind" (p.6). Therefore, Skehan (1998) and Long (2015) agreed that a task is a goal-oriented activity involving the real-world process of language use through a meaningful activity that engages various language skills and cognitive processes rather than focussing solely on the learning of specific linguistic features. In addition, task is seen as a unit of analysis throughout the design, implementation and evaluation of an analytic approach which is referred to as task-based language teaching (TBLT) (Long, 2015).

However, Ellis (2003, pp. 9-11) proposed that a task:

- is a workplan
- involves primary focus on meaning
- involves real-world processes of language use
- involves some kind of gap (e.g., reasoning gap)
- encourages students to rely on their own linguistic and non-linguistic resources
- can involve any of the four language skills
- engages cognitive processes
- has a clearly defined communicative outcome.

According to Ellis' definition, the syllabus, methodology and assessment are based on a synthetic approach which he referred to as task-supported language teaching (TSLT), as distinct from TBLT (Ellis, 2003, p. 65). The difference between both TBLT and TSLT will be discussed in the following section (see Section 2.2.2).

Despite the definition of task given above (i.e. as defined by Skehan and Long), Ellis (2017) argued that any task performed in the classroom ends up being 'pedagogic' rather than 'real-world'. Both of these types of tasks are primarily intended to deliver authenticity in the classroom. However, it is important to distinguish between the two as they comprise two different types of authenticity, namely situational authenticity (real-world tasks) and interactional authenticity (pedagogic tasks) (Bygate, 2016; Ellis, 2009c, 2017).

Target tasks or real-world tasks, as defined by Long (2015), are an "end-state, product-oriented, expert view of what the learner needs to be able to do" (p. 179). They refer to the language usage that is taken from the outside world which students will have to master in order to complete the course (Bygate, 2016). The main purpose of these type of tasks is to provide situational authenticity in such a way that the tasks are genuine and based on a real-life situation which occurs outside the classroom (Bygate, 2016; Ellis, 2017). Examples of target tasks or real-world tasks include a telephone conversation, taking an order from a customer, giving directions to a tourist, making a hotel room reservation etc. (Ellis, 2018; Long, 2015). In other words, 'task' is an activity that people do in their everyday life whether at work, for leisure and in between (Long, 2015). Therefore, the nature of these tasks is to encourage students to produce the authentic language which should occur in such situations.

Pedagogic tasks, in contrast, are derived from target tasks (Long, 2015) and they refer to tasks that are specifically designed to cater to the classroom's needs, although they may feature real-world contexts in some way. Pedagogic tasks adopt the learner's perspective to suit their level

of proficiency, usually in a "series of increasing complexity" and then gradually convert to target tasks (Long, 2015, p. 179). However, the aim is to provide interactional authenticity, even though they lack real-world language authenticity (Ellis, 2009c, 2017). An example is spot-the-difference tasks; it is very unlikely that, in a real-life context, people would engage in a conversation focussed on identifying the differences between two pictures. Although situational authenticity is absent in the practice, this type of task encourages cognitive processes, turn-taking and negotiations of meaning that reflect everyday language use and thus achieve interactional authenticity (Bygate, 2016; Ellis, 2017, 2018). For the purpose of the present study, Ellis's (2003) definition was adopted for tasks; all speaking tasks in the intervention used pedagogic tasks that included interactional authenticity that simulated real-world events and encouraged students to engage in cognitive processes in order to solve problems.

2.2.2 Strong and Weak Versions of TBLT

As there are many ways in which tasks can be employed in a language syllabus, advocates of TBLT have distinguished between TBLT and task-supported language teaching (henceforth TSLT) (Bygate, 2016; Ellis, 2003, 2018; Long, 2015, 2016; Skehan, 1996). In both TBLT and TSLT, communicative tasks are used to reinforce the students' use of the target language (Bygate, 2016) and designed to create opportunities for using language under real-world conditions (Ellis, 2018). However, the functions of tasks differ between the two contexts. The difference between the two types is that, in TBLT, tasks constitute the fundamental structure of the curriculum, whereas tasks in TSLT are not the defining unit of the curriculum or course (Bygate, 2016). Moreover, TSLT attempts to integrate explicit instruction within tasks based on a selection of linguistic features in a more complex pedagogic context (Skehan, 1996).

In TSLT, the teacher provides explicit instruction on the target features of the language to the students prior to the tasks (S. Li, Ellis, & Zhu, 2016). However, tasks in TBLT are conducted by the students freely without any explicit instruction or language input from the teacher. Moreover, TSLT does not serve as a unit for designing courses but is only considered as a methodological device for practising a certain target structure (i.e., free production) which aims to provide students with opportunities to demonstrate the correct use of the target language features while trying to achieve a communicative outcome (Ellis, 2018). Therefore, TSLT complements and supports conventional methods by extending the existing language-focused syllabus (Bygate, 2016). In TSLT, tasks are employed as communicative tasks to offer a free practice stage for students to use specific language patterns that they have been previously taught (Ellis, 2003). Therefore, Ellis (2018) claimed that TSLT is the weak version of TBLT and is compatible with the traditional presentation-practice-production paradigm because of its nature as product-oriented which is geared towards accuracy. This is because, the role of the task in TSLT is reduced to give students more opportunity to practice the language taught within the tasks and not the opposite, hence students are directed to attend to predetermined target forms and structures (Ellis, 2018). TSLT is driven by skill-learning theory where learning takes place when students develop their declarative knowledge of target features and then gradually automatize this knowledge through practice (DeKeyser, 1996, 1998). Moreover, TSLT is considered practical, likely to be acceptable to teachers (Ellis, 2003) and can be used alongside other conventional teaching methods (Shehadeh, 2005). However, TSLT is not to be confused with the traditional method, presentation-production-practice (PPP). The standard PPP is seen as teacher-centred and students are exposed to several ‘controlled activities’ such as simplified dialogues, reading passages, intensive written exercises via drills, followed by gradually ‘freer practice’ (i.e. pseudo-communicative language use) (Long, 2015). In addition,

in the PPP method, L2 is not considered as the object of instruction, whereas TSLT it is(Long, 2015).

Some advocates of TBLT have rejected TSLT and argued that it is similar to the context of the traditional method which emphasizes the importance of focus on forms (Long, 2015; Skehan, 1996). Focus on forms (FonFS) instruction emphasises the linguistic features of the language, which are often presented through distinct grammar rules and/or other metalinguistic information (Long & Robinson, 1998). Long (2015) rejected the idea of providing students with explicit instruction as it draws students' attention more towards the linguistic features (grammatical, lexical, phonological aspects etc.). Therefore, it reduces the role of the tasks and refutes the core purposes of TBLT (Long, 2015). In other words, TBLT is considered to be more effective as it is more process-oriented and tasks are designed to promote fluency. In addition, within TBLT, focus on form should be reactive and occur online while students are performing the tasks, rather than being proactive (i.e. pre-planned) (Long, 2016). Simply put, in TBLT attention to particular grammar structures arises incidentally and reactively, for instance when a learner has difficult using a particular structure during task performance. Similarly, Skehan (1996) states that tasks in TBLT should be the unit of the language teaching and the basis for the entire curriculum, and everything that comes with it, is considered supplementary. Moreover, the curriculum and pedagogic procedures within the curriculum are created primarily around the tasks rather than by selection of language priorities (Bygate, 2016). Therefore, TBLT serves as a mean for defining the content of the instructional course with the purpose of providing the students with opportunities to use the language naturally in order to achieve particular communicative outcomes (Ellis, 2018).

Based on the above discussion, both TBLT and TSLT share similar methodology (i.e. tasks) but the degree to which this methodology is applied may differ considerably. Tasks in TBLT are communication-focused in which the linguistic features are learnt incidentally, whereas

tasks in TSLT are used as vehicle for providing focussed practice of specific linguistic features (e.g. grammar and lexical items). As for the present study, both approaches were employed in two experimental groups. The pure TBLT approach was implemented in the ST (i.e. using speaking tasks only) group which was aligned with a strong version of TBLT (J. Willis, 1996), whereas the TSLT was employed in the EIST (i.e. explicit instruction prior to task and followed by speaking tasks) group which was associated with the weaker version of TBLT (Ellis, 2003).

2.3 Speaking and the Production of Speech

Speaking is a productive skill which involves receiving and processing information and producing systematic verbal utterances to convey meaning that happens in a 'real time' situation (Bailey, 2006), for example spontaneous conversation with another interlocutor. This process occurs spontaneously in the first language (L1). However, in an environment where English is a second language (L2), the skill of speaking is not automatically transferable from the speaker's first language to the second (Thornbury, 2005). A student who is proficient using the L1 may not be proficient in speaking the L2 as both languages have different knowledge and skills. In the following discussion, the speech production model for L1 by Levelt (1989) and L2 by Kormos (2006) are drawn upon and linked to the dimensions of speaking, namely complexity, accuracy and fluency, in order to give an overview of why speaking can be a very intricate skill.

2.3.1 A Psycholinguistics Model of Speech Production

The speech production model (De Bot, 1992; Kormos, 2006; Levelt, 1989) plays a major role in understanding the processes involved in students' speaking performance. Levelt (1989) proposed the 'speaking' model that outlines a framework of a number of information processing components during the process of speech. Later, it was adapted to the bilingual speech model by De Bot (1992) and Kormos (2006). The main characteristics of these model

are that they are "incremental and parallel, and the lower level processing is more automatized than higher-level processing" (De Bot, 1992, p. 1). In addition, these models (De Bot, 1992; Kormos, 2006; Levelt, 1989) involve similar basic psycholinguistic mechanisms in which they highlight the complex nature of speech production by describing how speakers encode or formulate ideas into speech (Kormos, 2011). Although these models have the same components (e.g. conceptualisation, formulation, articulation and monitoring), the processes that occur in L1 and L2 differ especially in relation to formulation (i.e. grammatical and phonological encoding).

In this study, the L1 speech production model by Levelt (1989) will first be explained in order to understand how L2 production works followed by the L2 model by Kormos (2006) as the L1 model forms the basis of the L2 model. The bilingual production model by De Bot (1992) is not discussed as it is concerned with code-switching which is beyond the scope of this study. The models provide researchers with interpretations on how speakers plan, process and produce language. The aspect of speech production which includes complexity, accuracy and fluency will then be discussed in the following section in relation to the speech models. The underlying constructs of attention and memory will also be presented.

2.3.1.1 Levelt's (1989) Monolingual Speech Model

Levelt (1989) proposed the psycholinguistic model of the process of speaking that involves several processes and sources of knowledge and skill. Levelt's model, which is used in this study, consists of four main components which are responsible for different processes of speech production. As illustrated in Figure 2.1 below, these components involve: 1) *conceptualisation*, planning what to say; 2) *formulation*, encoding information linguistically; 3) *articulation*, producing speech sounds; and, 4) *monitoring*, checking and correcting the produced output.

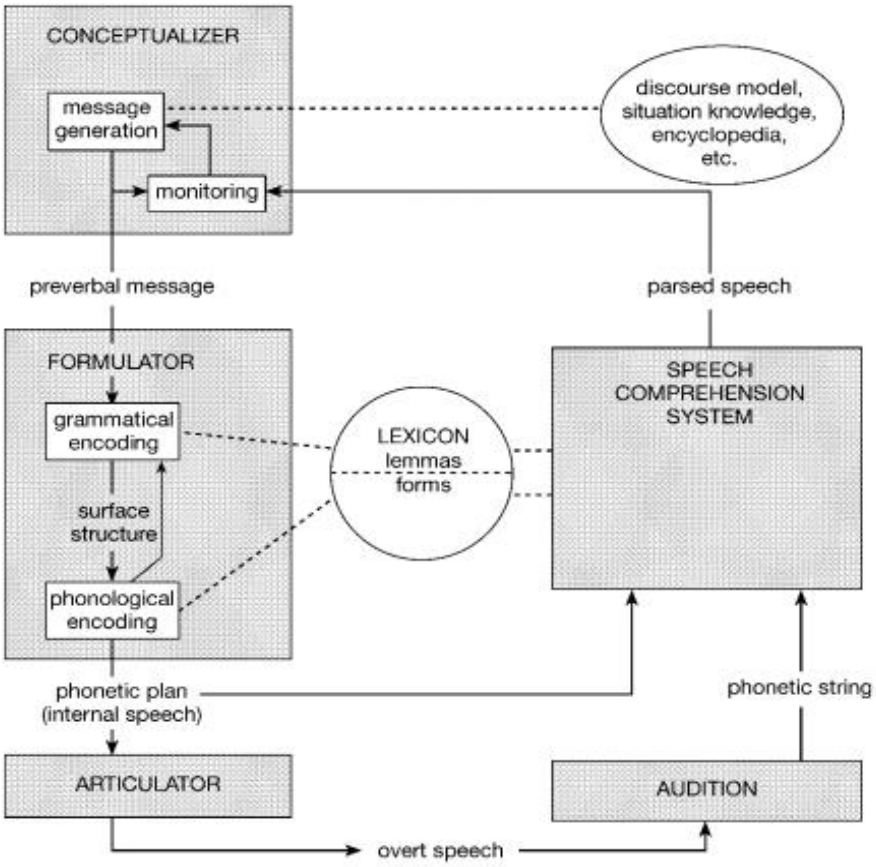

Figure 2. 1: Model of monolingual speech production (Levelt, 1989, p. 9)

Conceptualisation

Conceptualisation is considered to be pre-linguistic (Warren, 2013), in other words, it does not involve any form of language. Before speech is uttered, the conceptualiser starts to generate notions or abstract ideas without producing words or sentences. In this stage, two types of planning are involved, namely macroplanning and microplanning. In macroplanning, the speaker decides how to achieve their intended communicative goal by relying on his/her background knowledge and planning to use relevant speech acts (e.g. asking information, making a promise, giving route directions etc.). Microplanning involves the speaker performing detailed planning of each individual speech act. Once a speaker has made initial decisions on the communicative goal, individual speech acts can be planned in more detail.

The output of this planning is a preverbal message, which is not yet organised into the form of linguistic structures. This preverbal message becomes the input for the formulator.

In the classroom context, for example, the teacher introduces a topic on animals to the classroom. Students retrieve the idea that *'eagles fly'*. Information is retrieved on the lexical entries for *eagle* and *fly* in order to encode the sentence grammatically. At the same time, the students also retrieve additional information based on their background knowledge to the concept of the *eagle* (e.g. big, bird, predator, strong wings, sharp claws).

Formulation

A speaker needs to put together the elements of language that will express ideas within the preverbal message received from the conceptualiser by drawing on his/her own knowledge of the language, including grammar and vocabulary. In the formulation process, the mental lexicon provides *lemmas*, that contain semantic and syntactic information; and *lexemes*, which carry morphological and phonological specifications. During this process, the formulator generates a phonetic plan that contains two processes, grammatical and phonological encoding. In the grammatical encoding, the lemmas are accessed through a process of matching the meaning of the preverbal message with the semantic information which then activates the syntactic building procedures. Meanwhile, the phonological encoding accesses the morpho-phonological information stored in the lexemes that later produces a specific phonetic plan, which is transferred to the articulator. Following the example above, students make grammatical and phonological encodings as shown in Table 2.1 below.

Table 2. 1: Example of grammatical and phonological encodings

Grammatical Encoding		
Category	*Eagle*	*Fly*
Word class	Noun	Verb
Grammatical function	Countable	Subject + Verb (SV) = Eagles fly
Morphology	Regular plural	Flies, flying, flew, flown

Phonological Encoding	
Retrieving the forms	eagle + fly
Inserting phonological forms into structure	/*iː/ +/g*(ə)*l/ + /flaɪ/*
Adding the morpho-phonemic details	eagles + fly */iːglz/ + /flaɪ/*

Articulation

Producing the utterance involves the articulation process and the speech apparatus. In the articulation process, the phonetic plan is temporarily stored in the articulatory buffer. It is this buffered information where a speaker experiences internal speech. Then the articulator retrieves chunks of internal speech from the buffer and makes them available to be executed. The final outcome of this is overt speech produced by the articulatory organs.

Audition, Speech Comprehension and Monitoring

The overt speech is received by the audition component, where it recognises the articulated words. The speech comprehension system retrieves and records their meaning in working memory. In other words, the speech comprehension system scans both lemmas and lexemes, recognises words, retrieves meanings and parses the incoming speech. Thus, the speaker can monitor both the meaning and the well-formedness of his/her production. This is an important process called self-monitoring, located in the conceptualiser which is responsible for the outcome of the speech production processes. The monitor screens and compares the preverbal message with the original intention of the speaker before it is sent to the formulator. During this phase, the preverbal message might need some modifications if the speaker finds that the formulated message is not appropriate in terms of its content and communicative situation. When a problem occurs, the speaker will either ignore it, revise the preverbal message or generate a new one.

2.3.1.2 Kormos' (2006) Bilingual Speech Model

Most aspects of L2 production can be explained using the L1 speech production model as both models are similar in terms of the four main processes. However, there are some particularities in most aspects of L2 speech that require further explanation. One of the major differences is the degree of automaticity and the formulation process (Poulisse, 1997). Therefore, Kormos (2006) proposes the bilingual speech production model, shown in Figure 2.2 below. This model, which is based on Levelt's monolingual speech model, shares the same three knowledge stores: 1) knowledge of the internal and external world or *episodic memory*; 2) *semantic memory* including the mental lexicon which is assumed to have hierarchical structure consisting of three levels, *conceptual*, *lemma* and *lexeme*; and, 3) *syllabary* which stores the the syllables as chunks (Levelt, 1989). All knowledge stores are mutual between L1 and L2. That is, there is a common episodic and semantic memory for L1 and L2, a shared store for L1 and L2 lemmas and lexemes, and for L1 and L2 articulatory scores. However, in L2 production, Kormos (2006) argued that there is the existence of the fourth store for L2 specific knowledge, containing declarative memory of syntactic and phonological rules in L2.

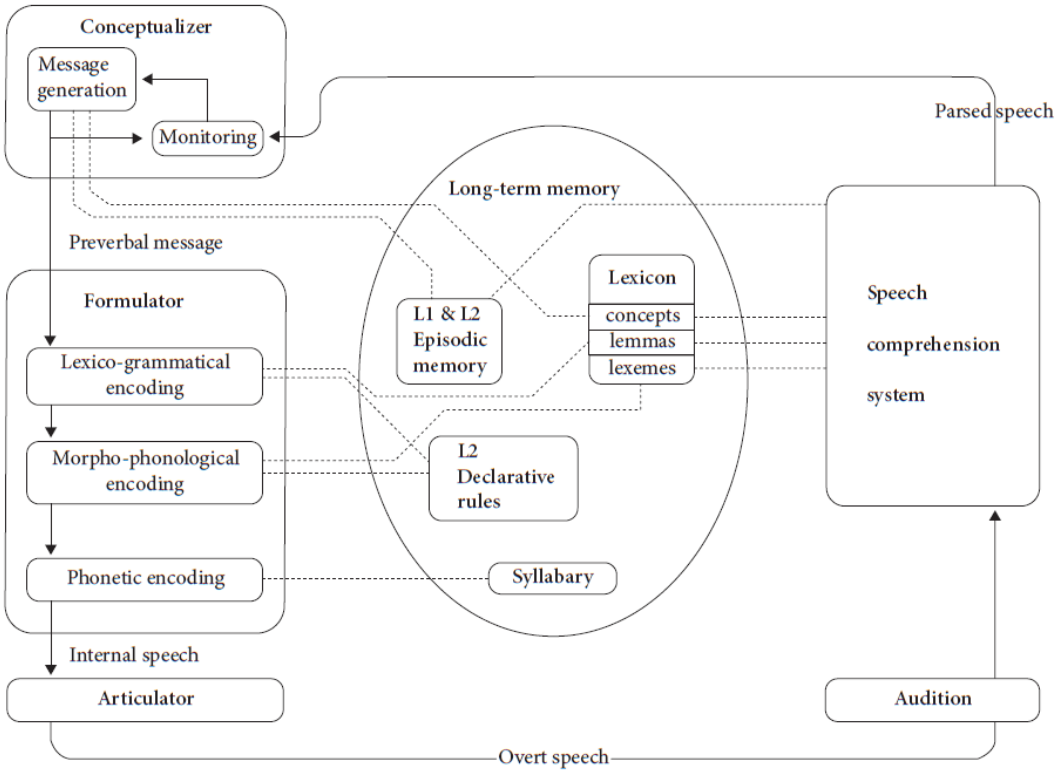

Figure 2. 2: The model of bilingual speech production (Kormos, 2006, p. 168)

In L1 speech production, message planning involves attention, whereas the process of formulation and articulation are automatic and effortless, and thus processing works in parallel (Kormos, 2011). This process refers to automatic processing where there is an absence of attentional control in performing cognitive activities (Kahneman, 1973). This involves several characteristics such as efficiency, rapidity, effortlessness, unconscious and uncontrollable nature (Kormos, 2006; Segalowitz, 2003). Generally, a person produces two to three words in one second and these words are selected from the thousands of words stored in the mental lexicon (Pollard, 2012). Therefore, no time is given in deciding what to say, because the formulation and articulation processes are automatic and function without awareness. This parallel processing of ideas and articulation can be accomplished easily because the specific morpho-phonological encoding draws on automatized linguistic knowledge that can be assessed and executed rapidly with the aid of ready-made language chunks (Tavakoli & Foster, 2008).

Nevertheless, such easy, automatized processing does not occur for the L2 speaker (e.g. particularly the less advanced speaker) because their syntactic and phonological encoding is underdeveloped and may not be fully automatized (Kormos, 2006). Therefore, the speaker's conceptualiser, formulator and articulator compete for limited attentional resources which subsequently result in slow speech or even silence (Tavakoli & Foster, 2008) and the production of erroneous utterances (e.g. ungrammatical structures or wrong choices of lexical items). The distinguishable features between L1 and L2 speech production can be summarised as follows:

- Syntactic and morphological rules are applied using declarative knowledge in L2 over procedural knowledge in L1.
- L2 speech production is controlled and the degree of automaticity may vary depending on the speaker's proficiency level.
- Lexical retrieval and encoding may require serial processing for L2 speakers which do not occur in parallel.

Kormos (2011) suggested the "need for a modified and extended model of L2 speech production in order to account for some of these differences in L1 and L2 production, namely the difference between L1 and L2 knowledge and the different nature of some L2 processing mechanisms such as syntactic and phonological encoding" (p. 42). Therefore, researching how L2 speakers process their speech in terms of its complexity, accuracy and fluency (see Section 2.3.2) offers valuable insight that enables understanding of the multifaceted nature of L2 processing.

2.3.2 Dimensions of Speaking

Speaking proficiency is part of the language curriculum and is an important object of assessment. Many studies evaluating speaking performance have drawn upon three main measures of speaking namely complexity, accuracy and fluency (henceforth CAF). Speech production should be assessed in terms of meaning (i.e. fluency) and form which is differentiated with regard to control (i.e. accuracy) and restructuring (i.e. complexity) (Skehan, 1998). These three components give effective indices for measuring and assessing students' speaking on a particular task (Skehan, 1998). However, these components are intricate, multidimensional and many researchers' interpretations differ on how these components should be defined and operationalized (Bulte & Housen, 2012; Bulte, Housen, Pierrard, & Van Daele, 2008; Housen & Kuiken, 2009; Norris & Ortega, 2009; Varcelotti, 2012).

2.3.2.1 Complexity

As the name itself suggests, complexity is the most ambiguous (Housen & Kuiken, 2009) and the most challenging construct within the CAF triad because of its "polysemous nature" (Palloti, 2009, p. 592). Generally, complexity can be defined as elaborated language (Ellis & Barkhuizen, 2005) and it involves students' mental effort and resources to internalise language features (Bulte & Housen, 2012). In other words, complexity can be described relative to proficiency as it challenges a student's interlanguage system that is not fully automatized to use a variety of sophisticated structures and vocabulary in the L2 (Ellis & Barkhuizen, 2005). Bulte and Housen (2012) proposed two components of complexity: grammatical complexity and lexical complexity. These components are generally interpreted as "the size, elaborateness, richness, and diversity of the student's linguistic L2 system" (Housen & Kuiken, 2009, p. 464). The construct of both grammatical and lexical complexity can be measured at three different levels (theoretical, observational, operational), as shown in Figure 2.3 and 2.4. The distinction between these three different levels is necessary for meaningful research interpretations to be

made. That is, it is necessary to recognise first what complexity is (theoretical), how can it be established in real language performance (observational) and how these behavioural manifestations can be measured (operational) (Bulte & Housen, 2012; Bulte et al., 2008).

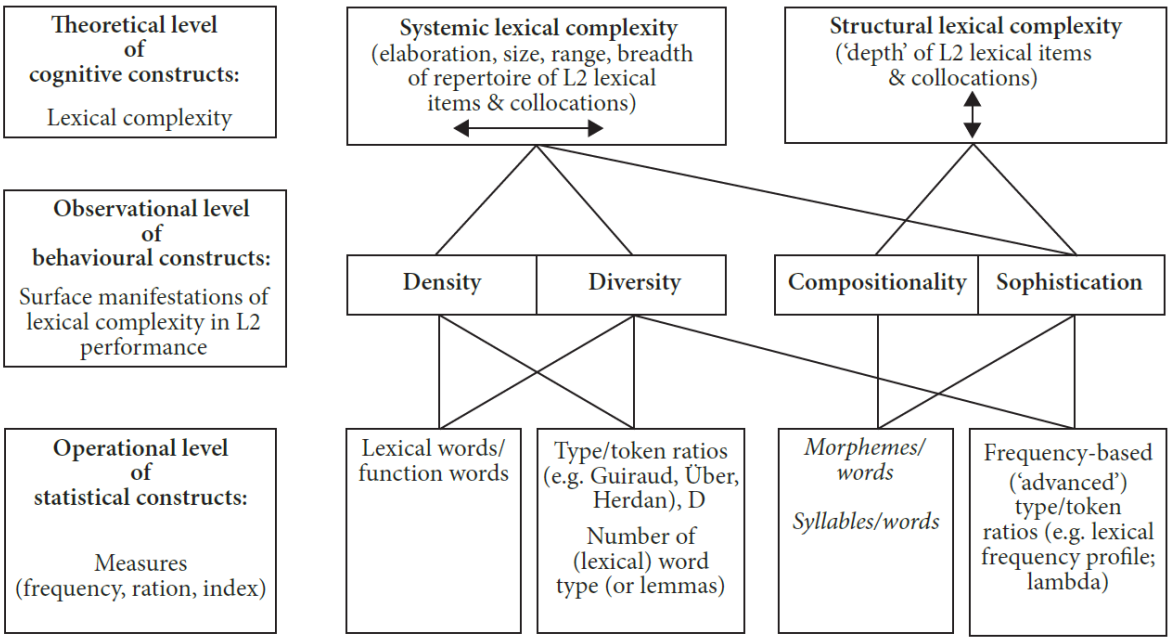

Figure 2. 3: Framework of lexical complexity (Bulte & Housen, 2012, p. 28)

In Figure 2.3 above, Bulte and Housen (2012) suggested that lexical complexity can be measured in terms of four different aspects of lexical complexity: density, diversity, compositionality and sophistication. Density measures the number of lexical words or content words (e.g. nouns, verbs, adjectives and adverbs) and function words or words that express a grammatical purpose (e.g. articles, prepositions, conjunctions etc.). Diversity refers to the learner's lexical proficiency by producing a variety of words which can be measured using *Guiraud Index*, *D*, etc. (Bulte et al., 2008). Compositionality assesses "the number of formal and semantic components of lexical items (e.g. phonemes, morphemes denotations)" (Bulte & Housen, 2012, p. 28). Lastly, sophistication deals with learner's knowledge of semantic relations (e.g. antonymy, synonymy, hyponoymy and hypernymy) (Bulte et al., 2008).

In assessing lexical complexity, several studies argued that there must be a distinction between lexical knowledge and grammatical knowledge, hence the need of lemmatisation (Courtney, 2014; Jarvis, 2002; Treffers-Daller, 2013; Treffers-Daller, Parslow, & Williams, 2016; Vermeer, 2000, 2004). This is because including inflections could create confusion between the students' vocabulary knowledge and grammatical knowledge (Jarvis, 2002). Vermeer (2004) argued that inflected forms would mask the learners' lexical knowledge and hence lemmatisation would be suitable of counting lexical richness. Treffers-Daller et al. (2016) investigated how different forms of lemmatisation impacted the measure of lexical diversity. The study involved 179 adult learners of English from 47 different countries which took part in the essay writing of Pearson Test of English (PTE) - Academic. As the word count in each essay differs from one another, a cut-off point of 200 tokens was used to treat all data equally. The essays were analysed in three ways: 1) lemma 0, no lemmatisation is involved; 2) lemma 1, lemmatisation, by using lemmas only and disregarding inflected forms of verbs, nouns and adjectives (e.g. treating the words *work, working, works* and *worked* as *work*); and, 3) lemma 2, by using the word family in which all inflected forms and derived forms were treated as the same type (e.g. *works, worked, worker*, *working*, and *workable* as *work)*. The results showed that lemmatisation had a significant effect on all lexical diversity measures (*TTR, Guiraud, D, HD-D* and *MTLD*). The scores were significantly lower in all measures if lemmatisations (lemma 1 and lemma 2) were applied than no lemmatisation. In addition, the results showed that most measures' (except *HD-D* and *MLTD*) scores from lemma 1 were higher than those in lemma 2 data. The values of eta square showed that lemma 1 explained most of the variance in scores at the different CEFR levels and were found to be more useful to disambiguate students' CEFR level than using lemma 2 or lemma 0 (Treffers-Daller et al., 2016). Therefore, the study suggested that using lemma (in line with lemma 1 above) is essential for measuring students' vocabulary knowledge and that lemmatisation should not remove derivational affixes. In other

words, a high level of lemma indicates that students are able to produce a wide range of vocabulary. As for the present study, lemmatisation was considered and employed in the complexity analysis in line with lemma 1 used in Treffers-Daller et al. (2016) (see Section 4.3 in Chapter 4 for details on the process of the lemmatisation and the data treatment).

Besides lexical complexity, Figure 2.4 below shows different sub-dimensions of linguistic complexity which can be evaluated across two main domains (e.g. syntax, morphology) and further analysed into smaller and fine-grained sub-domains.

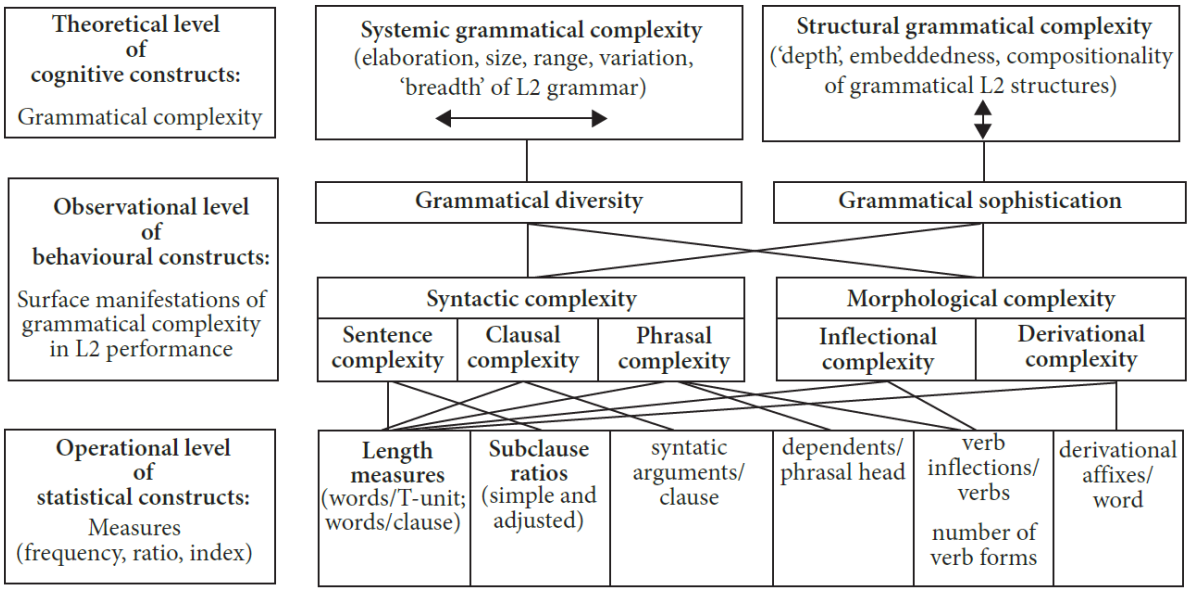

Figure 2. 4: Framework of grammatical complexity (Bulte & Housen, 2012, p. 27)

As shown in Figure 2.3 and 2.4, one measure is not necessarily associated with one particular construct (i.e. typically not a one-to-one match). In fact, one measure can have several purposes, demonstrated by the multiple lines starting from the lower levels. This is because different types of complexity measures are based on distinct construct of theory (Bulte & Housen, 2012). In reality, in using the language, Bulte and Housen (2012) described that "several of these complexity constructs may be closely intertwined which complicates their identification and assessment... the model presented here is merely a taxonomy and not a theory of complexity" (p. 26). Thus, this present study focused on the final tier of the framework

which contains what Bulte et al. (2008) describe as “the behavioural manifestations of the underlying cognitive constructs in actual L2 performance (production and reception)” (p. 279). Therefore, it could give a concrete indication of the degree of complexity of the student’s language proficiency that can be analysed more objectively (see Section 4.4.1 for details of the complexity measures used in the present study).

2.3.2.2 Accuracy

Accuracy is one of the oldest and the most explicit aspects of the triad (Housen, Kuiken, & Vedder, 2012). It can simply be defined as the ability to produce error-free speech (Housen &Kuiken, 2014). Ellis (2018) described accuracy as target language norms. It could easily be associated with the degree of language correctness versus erroneousness. In speech, Brown and Lee (2015) defined the term accuracy as articulately, grammatically and phonologically correct. These definitions “vary to some extent even in standard varieties of the target language and much more in the non-standard dialects that serve as the reference grammars for many L2 students” (Ellis, 2018, p. 141). Moreover, Foster and Wigglesworth (2016) argued that assessing students’ errors is a more reliable measure of accuracy which needs to be taken into account among L2 researchers. Palloti (2009) distinguished between the acceptability and the adequacy of speech production. That is, grammatically accurate language can be entirely inadequate for successful communication, whereas erroneous language can convey the meaning perfectly. Therefore, there must be clarity on what accuracy actually means depending on what is being researched and how it is operationalised (see Section 4.4.2 for details of the accuracy measures used in the present study).

2.3.2.3 Fluency

Fluency is the ability to produce speech in real time without having any interruption (e.g., pausing or hesitation) (Ellis & Barkhuizen, 2005). In simple terms, fluency can be described as speech that flows smoothly and naturally (Brown & Lee, 2015). This could imply that the

aim of fluency is to match the target language performance of a native-like behaviour (Palloti, 2009). Similarly to complexity, fluency is considered to be a multi-compositional construct (Ellis, 2018), thus it could be a challenge to determine which quantifiable linguistic phenomena contribute to the perception of fluency in L2 speech (Housen & Kuiken, 2009). Skehan (2009) distinguished three dimensions of fluency: 1) speed fluency (speech rate at which is performed); 2) breakdown fluency (pauses and silences that hinder the flow of the speech); and 3) repair fluency (hesitations and reformulations).

Tavakoli (2011) distinguished the pausing patterns between native speakers and L2 learners which greatly differed. Two groups of 40 English native speakers in their undergraduate and 40 L2 learners of English were recruited. The students' oral performances were elicited by four oral narrative tasks which were conducted in a monologic manner. The results showed that the pausing patterns differed between L2 learners and native speakers in terms of the number of pauses in their performance; L2 learners paused more frequently and produced more silence mid-clause. The study suggested that L2 learners were searching for a word, a structure or to think and plan their speech, whereas native speakers paused at the end of a clause to allow breathing space. This indicated that greater information processing load that occurred for L2 learners was associated with L2 acquisition. Segalowitz (2010) suggested that to master an L2 and achieve a high level of advanced level of speech; fluency requires automatization which enhances the fluidity and efficiency of speech in terms of the absence of dysfluency and pauses (see Section 4.4.3 for details of the measures used in the present study).

Drawing upon the discussion of CAF, Skehan (2018) characterised successful performance in task-based contexts as requiring "more advanced language, leading to *complexity*; a concern to avoid error, leading to higher *accuracy*; and a capacity to produce speech at normal rate and without interruption, resulting in greater *fluency*" (p. 35). In other words, complexity is achieved when the students are willing to take risks in attempting new and more complex

language structures; higher accuracy is seen when students' attention is focused on form and hence produce grammatically accurate language; and fluency is attained when students prioritise meaning over form. However, Skehan and Foster (2012) argued that to achieve such standards is a challenge and does not commonly occur for L2 learners, though it is not impossible by any means. Research within TBLT has studied whether these aspects of L2 performance compete against each other or complete each other during the process of speech production. Therefore, research in this domain is motivated by Skehan's (2009, 2014, 2015) Limited Attentional Capacity (LAC) which is explained in Section 2.5.2.

2.4 The Role of Explicit and Implicit Instruction and Knowledge in Second Language Acquisition (SLA)

The dichotomies of explicit/implicit instruction and explicit/implicit knowledge have consistently attracted the attention of researchers in the field of SLA (Ellis, 2016; Hulstijn, 2002; Long & Robinson, 1998). The present study looks into the development of students' speaking ability in terms of CAF (as discussed in the previous section) through the implementation of different forms of TBLT and not specifically at the types of knowledge that underpin these abilities. Nevertheless, it is useful to provide some contextual information about how these different types of knowledge are thought to develop. The following sections discuss in detail the dichotomies of implicit/explicit instruction and knowledge and the relationship between them.

2.4.1 Explicit and Implicit Instruction

When dealing with the high cognitive demands of a given task, students should be assisted with the language that they need to perform the task successfully (Skehan, 1996). That said, language instruction is considered as one of the many tools that can be used to assist students in achieving those task demands. Generally, language instruction is commonly used to assist

teachers and to intervene in the process of language learning (Ellis, 2018), with the purpose of helping learners overcome learning difficulties with particular aspects of the language. Numerous studies have compared two types of instruction, explicit and implicit instruction and their impact on language outcomes (Afsahi & Lotfi, 2016; Archer & Hughes, 2011; Ellis, 2009a, 2018; Hulstijn, 2015; Norris & Ortega, 2000). Implicit instruction aims to create an environment where students are able to learn "experientially through learning how to communicate in the L2" (Ellis, 2009a, p. 16). Implicit instruction is characterised as the absence of rules and activities in which meaning is emphasized (Hulstijin, 2005). Implicit instruction provides students with a flood of specific linguistic forms and patterns while students' attention is primarily drawn towards meaning and any attention to linguistic forms arises naturally without scaffolds or explicit teaching (Ellis, 2009a). Moreover, Long (2000) suggested that implicit instruction takes the form of focus on meaning (FonM) instruction, which emphasises incidental and unconscious learning where the students are entirely focused on meaning and less on grammatical correctness.

In contrast with implicit instruction, explicit instruction is characterised by a series of scaffolds in which students are guided through the learning process and given a demonstration of the target forms (Archer & Hughes, 2011). In other words, explicit instruction draws students' attention to what is to be learned and focuses more on a structural rather than a task-based syllabus. In explicit instruction, students are encouraged to develop metalinguistic awareness of the target forms (DeKeyser, 1995; Ellis, 2009a). Therefore, Long (1991) categorised explicit instruction as two approaches; focus on forms (FonFS) and focus on form (FonF). The former refers to the traditional language teaching which typically focuses on teaching discrete linguistic forms in separate lessons in a sequence predetermined by the teacher based on linguistic complexity (Laufer, 2006; Long, 1991, 2000). In detail, several criteria of FonFS have been conceptualised by Housen and Pierrard (2005, p. 10) as follows:

- directs attention to target form
- is *predetermined* and *planned* (e.g., as the main focus and goal of a teaching activity)
- is obtrusive (interruption of communicative meaning)
- presents target forms in isolation
- uses metalinguistic terminology (e.g., rule explanation)
- involves controlled practice of the target form

In contrast, FonF aims to draw students' attention to linguistic elements as they arise incidentally during communicative language activities (Laufer, 2006; Long, 1991, 2000; Long & Robinson, 1998). In other words, the basic principal of FonF requires the prerequisite of FonM before attention is given to linguistic elements (Doughty & Williams, 1998a). Following Long's definition of FonF, Ellis (2016, p. 407) conceptualised FonF as follows:

- arises in interaction involving the L2 student
- is reactive (i.e. occurs in response to a communication problem)
- is incidental (i.e. it is not pre-planned)
- is brief (i.e. it does not interfere with the primary focus on meaning)
- is typically implicit (e.g. it does not involve any metalinguistic explanation)
- induces 'noticing' (i.e. conscious attention to target linguistic forms)
- induces form–function mapping
- constitutes an 'approach' to teaching (i.e. FonF) that contrasts with a traditional form-centred approach (i.e. FonFS)

In sum, the difference between the two approaches has to do with how students see themselves working with the language. In the FonFS approach, language is viewed as the object of study and students see themselves as learners of a language, whereas for FonF, language is considered as a tool for communication and students see themselves as language users (Laufer,

2006). However, Doughty and Williams (1998b) argued that the term form must not be limited to grammar only but other aspects such as vocabulary should also be considered. This is because "it is likely that focus on form can enhance lexical acquisition. And there is mounting evidence that, in the acquisition of lexical items, as with that of grammatical structures, some interaction is helpful" (Doughty & Williams, 1998b, p. 212). Therefore, FonF in this present study can refer to both grammar and vocabulary acquisition.

According to the characteristics mentioned above, it can be concluded that explicit instruction in this study was achieved by implementing it to the students at the pre-task stage prior to the speaking tasks. The students that received explicit instruction (EIST group) were explicitly taught on the grammar and vocabulary items (i.e. definition of the target structure were given with examples via PowerPoint) related to the specific theme/topic followed by a short grammar/vocabulary exercise such as classifying, fill-in-the-blanks, close passage, error identification and sentence completion (see Section 3.5 for more detail on the implementation of the explicit instruction and the intervention).

2.4.2 Explicit and Implicit Knowledge

It is crucial to understand the meaning of explicit and implicit knowledge. In contrast to explicit and implicit instruction which focuses on a means of facilitating learning, the concepts of explicit and implicit knowledge are referred to as the product of learning (Rogers, Revesz, & Rebuschat, 2016).

A number of definitions have been proposed regarding implicit and explicit knowledge. Explicit knowledge is known as learned knowledge (Krashen, 1982) or declarative knowledge (DeKeyser & Criado, 2013) which is verbalisable and occurs consciously through controlled processing (Ellis, 2006). Krashen (1982) argued that explicit knowledge can be used when students monitor their language use, are focused on form, and given sufficient time to access

the knowledge. That said, explicit knowledge can also be developed when students encounter linguistic difficulty related to the L2 use (e.g. morphosyntactic and phonological rules). Within explicit knowledge, Ellis (2006) distinguished between analysed knowledge and metalinguistic explanation. The former refers to "conscious awareness of how a structural features works", whereas the latter is "knowledge of grammatical metalanguage and the ability to understand the explanation of rules" (Ellis, 2006, p. 95). Implicit knowledge, in contrast, is defined as procedural knowledge (DeKeyser & Criado, 2013) and characterised as unconscious and acquired through automatic processing outside of the students' awareness (Ellis, 2006). This type of knowledge is typically thought to underpin language use in more time-pressured, communicative contexts (e.g. fluent speech).

As for the relationship between explicit and implicit knowledge, it can be discussed in terms of three different positions: the strong, weak and non-interface positions (Ellis, 2006; Hulstijn, 2002; Krashen, 1981, 1982). The strong interface position hypothesised that declarative (explicit) and procedural (implicit) knowledge co-exist (Hulstijn, 2015). DeKeyser (1998) emphasises that explicit knowledge can be converted to implicit if students are exposed to and given the opportunity for extensive communicative practice. Drawing upon the widely held belief that 'practice makes perfect', this position is best exemplified in skill acquisition theory (DeKeyser, 2007) which suggests that language acquisition consists of and proceeds through "a series of sequenced stages, from initial representation of knowledge to highly skilled behaviour" (DeKeyser & Criado, 2013, p. 1). These stages involve the development of declarative knowledge, followed by proceduralization, and eventually automatization (DeKeyser, 2007; DeKeyser & Criado, 2013). In the declarative stage, the concept of 'knowing that' is emphasised, in which students accumulate an understanding about the language features, such as the grammatical rules. Then, the procedural stage can be described as 'knowing how', where students apply the declarative knowledge and perform particular

behaviours to solve a problem, thereby developing procedural knowledge of the target structure. Finally, automatization is the result of the restructuring of procedural knowledge; automatized knowledge can be drawn on spontaneously and effortlessly. Students are typically thought to have developed automatized knowledge once they have managed to reduce the time taken to use the language (e.g. they speak faster) and the error rate (i.e. the language produced become more accurate). Both the strong interface position and skill acquisition theory share the same basic principle, namely that conscious processes and thus declarative knowledge should be the starting point of learning (DeKeyser, 1998). Therefore, strong interface position would align with FonFS approaches which aim to facilitate learning through PPP (Dalili, 2011).

Unlike the strong interface position, which emphasizes the role of conscious knowledge and processes in SLA, the weak interface position differs depending on the weight it assigns to consciousness. The core tenet of these positions is that explicit and implicit knowledge plays an important role in learning, and that both types are possible outcomes of explicit or implicit instruction (Ellis, 1993; R. Ellis, 2005a; Ellis, 2006). In other words, students could gain awareness of structures they have been learning both explicitly and implicitly. Moreover, explicit knowledge draws the students' attention towards the target linguistic features and facilitates the development of implicit knowledge through *noticing* (i.e. attention to specific features in the input) and *noticing the gap* (i.e. comparing the targeted linguistic features in the input with their own existing grammatical knowledge) (R. Ellis, 2005a; Schmidt, 1990). Therefore, a position such as the weak interface position suggests the use of explicit instruction as a means of providing explicit information together with opportunities for language practice, which could facilitate the noticing process (R. Ellis, 2005a; Ellis, 2006, 2018).

In contrast, the non-interface position argues that not everything is learnable (i.e. with conscious attention), hence placing much less emphasis on the role of explicit instruction (Han & Finneran, 2013). The non-interface position argues that there is a clear distinction between

explicit and implicit knowledge, and that the two are entirely different acquisitional mechanisms (Hulstijn, 2002; Krashen, 1981). The two types of knowledge are processed differently; explicit knowledge is controlled, whereas implicit knowledge requires automatic processing (R. Ellis, 2005a). Krashen (1982) argued that acquisition (implicit learning) leads to the development of implicit knowledge or the acquisition of competence in which students are not aware of the grammar rules acquired but develop an instinct about the correctness of the language. Learning (i.e. explicit learning), in contrast, occurs when students learn the grammar and rules of the language and are able to talk about them, hence producing explicit knowledge (Krashen, 1982). In addition, knowledge that has been learned explicitly is "not deployable in real, spontaneous communication" (Han & Finneran, 2013, p. 4). Therefore, acquired knowledge and learned knowledge are dissimilar and distinguishable (R. Ellis, 2005a; Hulstijn, 2002; Krashen, 1981, 1982). A study by Paradis (1994) supported the claim of Krashen by suggesting that explicit and implicit memories are stored in different parts of the brain. Paradis (1994) argued that formal learning of the L2 cannot become procedural, whereas L1 competence is the results of acquisition and the L1 language process is automatic.

As for this study, the aim was not to examine the effect of two different forms of language teaching on the types of knowledge developed, but to investigate the differential impact of these two instructional forms (TBLT and TSLT) on the development of students' speaking performance. It is worth noting that TBLT and TSLT were designed to reflect implicit and explicit instruction, respectively. However, TBLT does not imply that students will learn the grammar and vocabulary implicitly nor develop implicit knowledge. This is because the characteristics of the speaking tasks (e.g. focused tasks) might help students (ST group) to learn the grammar and vocabulary items intentionally (see Section 3.5 on the characteristics of the speaking tasks used in the intervention).

2.5 The Role of Attention in SLA

The role of attention plays an important role in the L2 learning process, hence it is seen as a construct that have pervaded, explicitly or implicitly in SLA studies since its inception (Leow, 2015). Robinson (2003) argued that language learning and performance both require attention and subsequent encoding of information. In the present study, two constructs were drawn in which attention involves selection and capacity.

2.5.1 Attention as Selection

The key issue in studies of SLA is whether the attention to form that occurs during language learning which mainly focused on meaning, involves awareness. Krashen (1985) claimed that adult learners are still 'acquirers' and able to utilise the innate learning mechanisms involved in L1 acquisition; hence, conscious learning has minimal influence on the ability to learn and use the L2. However, Schmidt, (1990, 2001) through his Noticing Hypothesis, argued that conscious attention to the form within the input is crucial to subsequent L2 learning. Attention to input is needed in order to have the momentary subjective experience of 'noticing' the form that the learners are to subsequently learn (Schmidt, 1990). In other words, noticing arises when learners put their attentional resources to a specific aspect of the language (syntactic, morphological, phonological etc.) (Iwanaka & Takatsuka, 2007). If learners pay selective attention to a specific form, for example, it is likely that noticing of that form will occur. N. Ellis (2005) expressed that in order to establish a new form in long-term memory, conscious attention is needed. Most theoretical underpinnings in SLA are in line with Schmidt's Noticing Hypothesis and emphasise the vital role for learners' attention to form, whether aural or written, manipulated or authentic (Robinson, 1995; Tomlin & Villa, 1994; VanPatten, 2004).

Mackey (2006) claimed that if learners are consciously attending to the linguistic features, they are more probably to learn the L2. Therefore, noticing the form is likely dependant on the salience of the linguistic features such are grammar and the vocabulary items. For example, the use of suffix *–s* in English which can be used to show plurality or a third-person singular of the verb. The plural *–s* in English is one of the grammatical rules that is considered more "salient and communicatively functional in context that may be acquire implicitly" (Ellis, 2018, p. 77), whereas third-person *–s* can be less or non-noticeable and redundant, and may be acquired by learners if explicit guidance is given (Ellis, 2018). Therefore, this indicated that different types of focus on form are used to facilitate learners in learning the linguistic features of the L2.

As for the present study, the concept of noticing was adopted in relation to the EIST group, in that explicit instruction was given in the pre-task stage, whereas the ST group did not receive any explicit instruction. The EIST group was taught the definition and given examples of the grammar and vocabulary items prior to the speaking tasks. By providing explicit information and making students notice the target linguistic features, it was hypothesised that it would facilitate students' speaking performance (see Section 3.5 on details of the explicit instruction implemented in the EIST group).

2.5.2 Attention as Capacity

Studies of SLA show that attention is a limited set of mental resources which is shared across various processing activities (de Bot, 1996). Skehan (2009, 2014, 2015) proposed the Limited Attention Capacity (LAC) Hypothesis, known as the Trade-off Hypothesis. The hypothesis stated that learners' attentional resources are limited and that trade-offs between specific dimensions (e.g. complexity, accuracy and fluency) are "pervasive and unavoidable" (Skehan, 2018, p. 56). According to Robinson (1995), attention is seen as capacity-limited when the demands of the task are high, (e.g. complexity of the task, limited planning time etc.) thus these

constraints force learners to decide whether to focus their attentional resources on attending to lexical complexity, syntactic and/or accuracy (Skehan, 2015) especially for beginner learners who may not be able to attend to syntactic features and accuracy simultaneously.

Skehan and Foster (2012) suggested that there is a relationship between complexity and fluency and they are frequently connected, and likewise for accuracy and fluency. However, it is unlikely, though not impossible by any means, for both complexity and accuracy to be elevated at the same time (Skehan & Foster, 2012). Skehan argued that the trade-off effect between complexity-accuracy is inevitable as both are competing for the same pools of attention. This is because, in relation to the speech production model, complexity is generated by the conceptualiser, whereas accuracy (e.g. morphosyntactic building) is created by the formulator (Skehan, 2018). Unlike the L1, the conceptualisation and the formulation processes are different and do not occur in parallel in the L2 (Kormos, 2006). Hence, the speech production process becomes serial in nature when encountering both complexity-accuracy since "greater demands are made on limited attention resources" (Skehan, 2018, p. 59). Several studies have shown that different task demands or characteristics may influence the trade-off effects between complexity-accuracy such as different planning conditions (Foster & Skehan, 1996) and task structure (narrative telling) (Tavakoli & Skehan, 2005).

Therefore, this indicates that not all aspects of language could be increased simultaneously as certain tasks may enhance certain performance areas. Certain tasks may foster complexity and fluency, while other tasks advance accuracy and fluency. Theories of LAC are based on the assumptions that attention and noticing are essential for language acquisition (Schmidt, 1994) which can be assigned and reallocated within the stages of Levelt's (1989) monolingual speech production (Skehan, 2014).

2.6 Studies on TBLT versus TSLT

The debate regarding implementing explicit grammar instruction has been ongoing in the literature for a number of years (R. Ellis, 2005a; Ellis, 2006, 2018; Ellis, Li, & Zhu, 2019; Long, 1991, 2016). According to S. Li et al. (2016), linguistic guidance has its own benefits and is necessary in the pre-tasks stage of a lesson, that is to provide explicit instruction. Meta-analyses of form-focused instruction indicate that explicit instruction is more effective than implicit instruction (Norris & Ortega, 2000; Spada & Tomita, 2010). Norris and Ortega (2000) discovered that explicit instruction induced more substantial, target-oriented and durable gains than implicit instruction. Similarly, Spada and Tomita (2010) synthesised research into the effect of different types of instruction on the complexity of grammatical features, and the results showed greater gains from explicit instruction on simple and complex features than from implicit instruction. However, not all of the explicit treatments in the previous studies included in these meta-analyses were task-based (S. Li et al., 2016). Therefore, there is a need to investigate whether explicit instruction through tasks (i.e. TSLT) would benefit students' overall language outcome. This is because TBLT is primarily FonM and aims to develop students' language fluency, but the traditional method focuses on explicit teaching of the linguistic rules and students' language accuracy (Long, 2015), hence different types of instruction implemented in the classroom aim to develop and meet students' linguistics needs differently. By combining both instruction types, in which explicit instruction is combined with tasks, it is expected that students could develop both language fluency and accuracy simultaneously.

Some advocates of TBLT reject the idea of providing students with explicit instruction before a task (Long, 2000, 2015, 2016; Skehan, 1996, 2018). Long (2015) argued that "TBLT invokes a symbiotic combination of implicit and explicit learning" (p. 8). Ideally, attention to form should occur online when students are performing a task and not prior to the tasks (Long, 2000)

to enable “form-function mapping central to acquisition” (S. Li et al., 2016, p. 208). That said, reactive FonF is preferred in TBLT rather than the proactive FonFS. However, applying explicit instruction in a task-based scenario may ‘fill in the gaps’ between traditional synthetic syllabi and meaning-focused approaches (Ellis, 2018; Ellis et al., 2019; S. Li et al., 2016). Shehadeh (2005) favoured TSLT in the context of English as a foreign language (EFL) instruction as it helps students to learn the language correctly even with less exposure to the target language. Similarly, Sato (2010) argued that, in the context where the target language is foreign and not widely used (i.e. English language in Japan), tasks are suitable but best reserved for the free production stage of PPP approach because it allows students to experience meaning-form connections during language practice without relying heavily on controlled drills and exercises.

For this study, the main issue is to investigate whether explicit instruction followed by a speaking task (i.e. TSLT) results in better learning rather than simply performing the speaking task (TBLT) by itself. There were mixed results from several studies, which investigated the effectiveness of tasks in eliciting the target structure when explicit presentation of the structure was not provided. Mackey (1999) discovered that focused tasks (i.e. tasks that induce learners to produce a particular linguistic feature) are effective in developing more advanced structures and encouraging students to attempt and produce the target features. However, Samuda (2009) observed different results where students avoided using the target words (modal verbs) during task performance. However, only after students were guided by the teacher and taught explicitly about the target words did they start to utilise the target words in the task (Samuda, 2009). A study by Boston (2010) revealed that oral narrative tasks failed in encouraging students to use the target structure (passive voice) due to the fact that the task itself was poorly designed. The theme of the tasks was considered unsuitable to elicit use of the target structure, thus the active voice was frequently used (Boston, 2010). These shortcomings, therefore,

reflect that it is difficult to ensure that implementing focused tasks only is sufficient for students to use the target form or structure, let alone using unfocused tasks which are not designed for the use of a specific form (Ellis, 2018).

Other studies have investigated whether explicit instruction prior to a task encourages students to produce the target linguistic features. De la Fuente (2006) studied this question by exploring the impact of using explicit vocabulary teaching on students' vocabulary acquisition. Three groups following different approaches were formed. These approaches were the PPP group (n=10) which incorporated the traditional method; the TB-EF group (N = 10), incorporating explicit instruction in the post-task stage; and the TB-NEF group (N =10), task without explicit guidance from the teacher. 30 students, who were at an elementary level of Spanish, were assigned in three intact classes. Although students were taught Spanish in high school (1 to 2 years), they showed no previous knowledge of the 15 Spanish words that were going to be tested in the study. The procedures of the study were as follows: 1) students listened to a dialogue in which all 15 Spanish words were mentioned (provided in written form as well) and answered comprehension questions; 2) they listened to the second dialogue with the purpose of recycling functional and lexical aspects but with a different setting; and, 3) they were required to take an immediate vocabulary test and a delayed vocabulary test one week later. Then students worked in pairs and were asked to role-play the given dialogue. The analysis of the vocabulary tests involved reviewing the recorded audio and scoring students' correct vocabulary usage of the 15 Spanish words. One point was given when students produced the correct form. The statistical results of the study indicated that all groups failed to retrieve word forms in the delayed vocabulary test. Moreover, the study suggested that students in both task-based groups managed to negotiate the words as they were both noticing (i.e. due the salience of the words) and focusing on meaning, whereas the PPP group failed to do so. However, the

TB-EF approach was more effective in promoting acquisition of word morphological aspects as students in this group made more attempts in using the target words.

A study by Mochizuki and Ortega (2008) investigated whether pre-task planning with grammatical guidance enabled students to attend to a specific L2 form in the beginning-level of a foreign language classroom. First year high school Japanese students were asked to complete a pair-work storytelling task under one of the three conditions: without prior planning (*N*=17), with five minutes of unguided planning (*N*=20) or with five minutes of guided planning (*N*=19). For the procedure of this study, the non-planning group was asked to retell the story immediately after listening to an audio narrative stimulus. Meanwhile, both planning groups were given five minutes to prepare for the storytelling task but the guided-planning group received an additional handout which briefly explained relative clauses during the five-minute preparation time. The oral tasks were analysed based on global complexity by using the T-unit (e.g. the number of words, the number or subordinate clauses, and the number of relative clauses of which each of the measures were divided by the number of T-units of each narrative) and global fluency (e.g. speech rate and mean number of words per minute). The results of this study revealed that students who received guided planning produced twice the number of relative clauses and were more target-like in their oral narratives compared to the unguided-planning and non-planning group. However, for fluency, the unguided-planning group was found to be more fluent than their counterparts, in which they produced substantially longer narratives than the other two planning groups. This indicated that external guidance with planning facilitated students in their oral task and achieved task completion as their attention was oriented towards L2 form. Despite the positive effects achieved, the study suggested that planning time does not necessarily ensure higher accuracy of production.

Another study on planning, Ahmadian (2012), explored whether guided careful online planning has any effect on accuracy (English articles – a/an and the), complexity and fluency. The

participants of this study were 45 intermediate EFL students between the ages of 21 and 25, who had learned English for about 8-9 months. They were assigned into three different groups with each group containing 15 students: guided careful online planning (GCOP), unguided careful online planning (UCOP) and pressured online planning (POP). Students were required to watch a 10-minute episode of a silent classical film and later were asked to narrate the story. For both planning groups, students were allowed to take as much time as they needed, whereas the PO group was only given eight minutes for planning. The narrative tasks were analysed based on complexity by using the AS-unit (e.g. ratio of clauses, total number of different grammatical verb forms and mean length of AS-units), accuracy (correct usage of English articles) and fluency (number of syllables and number of meaningful syllables produced per minute of speech). The findings of this study revealed that the GCOP group produced English articles more accurately than the other two groups. As complexity, both guided and unguided online planning have positive effects on all measures of complexity compared to the pressured online planning. However, the POP group was found to be more fluent in their speech as students managed to produce more syllables per minute of speech than the groups which were given unlimited planning time. The study suggested that unguided careful online planning was more effective in producing more complex language and providing explicit guidance of the grammar items (English articles) does not affect complexity.

S. Li et al. (2016) examined the effectiveness of task-based and task-supported instruction in the acquisition of the English passive construction. A total of 150 Chinese middle school students were randomly assigned to form five groups: control, FonM, EI+Task (explicit instruction with task), TCF (task with corrective feedback) and EITCF (explicit instruction with task and corrective feedback). Students were given two dictagloss tasks with identical procedures and students were required to listen to the narrative presented by the teacher and later worked in pairs to retell it to the class. The study revealed that the explicit instruction had

a greater effect on students' accuracy especially the EITCF group which received both explicit instruction and recast feedback. However, the study suggested that these effects were largely related to the acquisition of explicit knowledge as none of the groups outperformed the control group on the elicited imitation test (EIT), which measured procedural/implicit knowledge. Explicit instruction had positive effects on students' accuracy but it failed to develop students' automated knowledge. The study also suggested that students with prior knowledge gained higher EIT scores than those who had none. However, the study did not report what effect different instruction had on students' production in terms of fluency and overall accuracy.

Van de Guchte et al. (2017) investigated the effects of different types of form-focused instruction during pre-task planning on students' oral task performance (accuracy and complexity). A total of 48 students, who were in the ninth-grade learning German as a foreign language, participated in the study. They were randomly allocated into two different planning conditions: focus on language (FonL) and focus on content (FonC). Students in both groups were required to watch two videos with each group focusing on two different aspects. The FonL (n=25) group observed the use of two-way prepositions with dative case, whereas the FonC (n=23) observed the appeal and persuasiveness of the presentations, both guided by instructions on a written handout. The transcriptions of the students' oral tasks were coded for the AS-unit or the Analysis Speech unit which 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, Tonkyn, & Wigglesworth, 2000, p. 365). The analysis of this study used the involved two measures of regarding attempted use of the target structure and the accurate use of the target structure, and three complexity measures (means of the total number of words per Analysis of Speech Unit or AS-unit; amount of subordination per AS-unit; and amount of coordination per AS-unit). The results showed that the FonL group used the target structure substantially more often and more accurately. For complexity, however, the

FonC used more coordinated clauses and subordinated clauses. Therefore, different types of form-focused instruction led to different language outcomes. The study suggested that focus on language may promote more and accurate use of the target structure, whereas focus on meaning elicit more complex task performance.

A more recent study by Ellis et al. (2019) examined the influence of pre-task explicit instruction on focused task oral production (complexity, accuracy and fluency). A total of 72 Chinese students in their eighth-grade were randomly assigned to two groups: task with explicit instruction (TEI) and task only (TO). Two narrative texts were used which each contained 15 sentences in the past passive form. Prior to that, the TEI group received 10 minutes explanation on past passive form and how to use it. The two texts were read aloud by the teacher three times, and students were required to listen and not to take notes. Later, students were given 15 minutes to work in pairs and then practise retelling the story. The narrative tasks were analysed using global measures which included complexity (length of AS-units), accuracy (errors per 100 words), and fluency (average pause length). In addition, the two additional measures were employed in this study which were target use and correct target use in order to assess the number of attempts in using passive voice and the number of correct used passive constructions, respectively. The study found out that there was no statistical difference in the two groups' accurate use of past passive form. However, thorough analysis showed that the TO group used more complex, more accurate and more fluent language in both tasks compared to the TEI group. However, the TEI group which had explicit representation of the target structure did not perform in all three measures (CAF) as compared to the TO group which was more complex, more accurate (global accuracy) and more fluent. Although the TEI group did not improve, the study suggested that explicit instruction led students to more attempts at using the target structure.

In summary, these six studies offer several interesting findings. First, focused tasks were effective in encouraging students to produce complex target structures in the task-only condition (Ellis et al., 2019; Van de Guchte et al., 2017). This could be caused by the nature of the tasks. In both Ellis et al. (2019) and Mochizuki and Ortega (2008), students listened to narratives read by the teacher and linguistic structures were presented by the power point slides. This gives the students the opportunity to comprehend better and activate their schematic knowledge. Second, explicit instruction encouraged students to use the target structures. This is not surprising as it is the main objective of TSLT and the importance of noticing the target structures. If students' awareness is drawn to specific structures, they will notice and use the structures in task performance. Therefore, students who have access to an explicit representation of a linguistic feature and the opportunity to practice this feature facilitate students in learning the L2 (Ellis, 2018). However, explicit instruction does not necessarily result in more accurate production of the target structure (Ellis et al., 2019) and consistent vocabulary retention (de la Fuente, 2006) although it did in some studies (Ahmadian, 2012; S. Li et al., 2016; Mochizuki & Ortega, 2008). Moreover, it is clear that explicit instruction might have impeded students' fluency based on the results showed in Ellis et al. (2019) and Van de Guchte et al. (2017). Students' attention was more focussed on producing the target structure, hence they failed to produce fluent language. This could be supported by the LAC Hypothesis (Skehan, 2009, 2015) in which students were not able to focus on all aspects of speaking simultaneously as their attentional capacity was weighted to one or two aspect(s) only.

The existing research to date is limited but suggests that, to some extent, focused tasks might not be enough for students to produce the target form, even when explicit instruction has been given. Moreover, these previous studies did not investigate lexis, complexity, accuracy and fluency in depth, as the whole concept of CAF. Some studies looked into CAF as a global measure (Ellis et al., 2019; S. Li et al., 2016), others looked into one or two dimensions of CAF

(de la Fuente, 2006; S. Li et al., 2016; Van de Guchte et al., 2017) and focused on certain grammatical structures (Ahmadian, 2012; Ellis et al., 2019). Therefore, the aim of this study is to investigate whether explicit instruction in the pre-task stage alongside tasks (TSLT) facilitates successful speaking outcomes in terms of both global and specific measures of complexity (lexical complexity, lexical productivity and syntactic complexity), accuracy (error-free speech) and fluency (numbers of breakdown and repair fluency). Further details on how explicit instruction through tasks is operationalised will be explained in Chapter Three.

2.7 Second language (L2) Motivation

The complexity of speaking in a second language may cause language learners to experience a high level of anxiety (Heng, Abdullah, & Yusof, 2012) and thus speaking is seen as an "anxiety-provoking skill" (Dincer & Yesilyurt, 2017, p. 3). Due to its difficulty and intricacy, it may cause language learners to feel anxious when speaking which could impact on their desire to persist in language learning and their level of motivation towards speaking. The theme of L2 motivation has captured the attention of many SLA researchers as it is one of the most important individual difference variables (Dörnyei, 1994, 2005; Gardner, 1985). This is because L2 motivation occupies a particular position that may affect one's success or failure in L2 learning (Dörnyei, 2005; Dörnyei & Kormos, 2000; Dörnyei & Ushioda, 2011; Gardner, 1985). Nevertheless, defining motivation has become a challenge to many researchers in the field of SLA as "L2 motivation is an eclectic, multifaceted construct" (Dörnyei, 1994, p. 279). Dörnyei and Ushioda (2011) described motivation as:

- why people decide to do something;
- how long they are willing to sustain the activity;
- how hard they are going to pursue it (p. 4)

According to Dörnyei and Ushioda (2011), research in L2 motivation is basically concerned with measuring the extent to which learners choose to undertake and persist with certain activities and the factors that influenced them which later prompted them to behave a certain way. Therefore, the importance of L2 motivation becomes essential, as learning a L2 such as English means not only learning the linguistic features, but also learning the cultural values and norms of the target language.

The nature of L2 motivation varies in different countries or environments. For example, learning English in a multilingual context in Malaysia might be affected by different motivational forces than learning modern languages (e.g. German, French) in an Anglophone context. In the former, English is seen as a second language which is used interchangeably (i.e. code-switching) with the Malay language; it is a compulsory subject taught in schools and universities; and is heavily used outside the formal settings. However, the latter, modern languages, are examples of foreign languages "almost totally absent from learners' environment beyond the educational context" (Krüsemann, 2018, p. 18). In addition, due to the multiracial and multilingual nature of Malaysian society, the new bilingual policy (see MoE, 2014) and the recent implementation of the National Education Blueprint (see MoE, 2013; MoE, 2015b) were introduced. Furthermore, with the exposure to different languages in a multicultural environment, learning English as a second language bridges the gap between different ethnic groups and act as a main force to enhance the emergence of intercultural communication (Cheng & Dörnyei, 2007; Clément, Dörnyei, & Noels, 1994; Dörnyei, 2005). Therefore, all of these factors may influence the nature of motivation in Malaysia.

Dörnyei (2005) indicated that L2 learning is a complex, lengthy and an ever-changing process when compared to other subjects, meaning it has to be sustained through motivation. That complexity has led to several theories which are discussed in the following section and are employed in the present study.

2.7.1 Gardner's Socio-Educational Model and Motivation Theory

The study of L2 motivation emerged from the pioneering social psychological perspective of social psychologists, Robert Gardner and Wallace Lambert, involving the social context of learning and attitudes between two different ethnolinguistic communities, in which the 'socio-educational model' was developed (Gardner, 1985; Gardner & Lambert, 1972). The development of the model began in Canada based on the evaluation of the "confrontational coexistence of the Francophone and Anglophone communities" (Dörnyei, 2005, p. 67) in which two official languages, English and French were used. It was believed that within this context, motivation to learn the L2 served as a bridge to keep the two communities together and as a main force responsible for enhancing intercultural communication (Cheng & Dörnyei, 2007; Dörnyei, 2005). The main principle of this model is that one's attitudes towards the L2 and the L2 community may influence learning behaviour towards the target language (Dörnyei & Ushioda, 2011). These attitudinal dimensions differentiate language learning from learning other subjects in which is it not solely an acquisition of knowledge or an accumulation of facts; language learning involves not only forms and structures of the language, but also understanding the culture values of target language community (Gardner & Lambert, 1972). Williams (1994) expands this further by stating that language learning "involves alteration of self-image, the adoption of new social and cultural behaviours and ways of being, and therefore has a significant impact on the social nature of the learner" (p. 77).

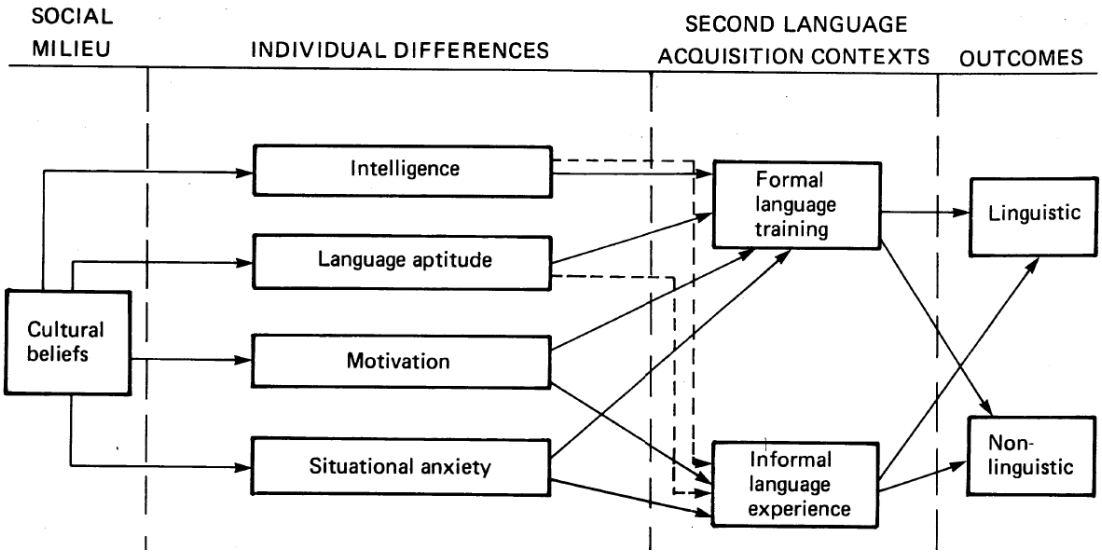

Figure 2. 5: The socio-educational model (Gardner, 1985, p. 147)

Gardner's socio-educational model shown in Figure 2.5 consists of four key variables: social milieu, individual differences, second language acquisition contexts and language outcomes. The process of language acquisition involves a causal interplay of these key variables (Gardner, 1985). Cultural beliefs derive from the learners' social milieu. This means the beliefs within the community of the learners would influence the learning process of the target language. Gardner (1985), for example, argued that if the community believes that learning the L2 is challenging and difficult, the overall level of language proficiency will be low and hence could interact with individual differences such as intelligence, language aptitude, motivation and situational anxiety. Gardner and MacIntyre (1992) classified these individual differences into two main categories: cognitive factors and affective factors and later added other variables within these factors. Cognitive factors include intelligence, language aptitude and language learning strategies, whereas affective factors comprise attitudes and motivation, language anxiety and self-confidence. These variables play different roles within the process of L2 learning and may affect learning outcomes differently depending on the socio-cultural context (Gardner & MacIntyre, 1992). In addition, these variables interact and function differently within formal and informal SLA contexts.

A formal language context refers to the situation where learners are exposed to direct instruction in the language (i.e. formal language classroom learning), whereas informal contexts occur outside classroom settings. Both cognitive and affective variables would be directly involved in a formal context. Gardner and MacIntyre (1992) proposed that cognitive variables play an important role in making the learning process easier while the affective variables lead learners to react to their learning environment. However, in informal contexts, motivation arguably plays the most significant role in language learning and cognitive variables are more secondary. Nevertheless, in both contexts, both types of variables determine learning outcomes to a certain extent, both linguistic (e.g. vocabulary knowledge, grammar, fluency etc.) and non-linguistic (e.g. attitudes towards the language, learning situation, values etc.). Positive experiences within any learning context are more likely to produce a high level of linguistic competency and improve levels of non-linguistic outcomes, and vice versa (Gardner & MacIntyre, 1992).

Regarding the individual difference of integrative motivation, Gardner's model includes three main components:

- Integrativeness: includes integrative orientation, interest in the target language and reflecting one's willingness towards the L2 community.
- Attitudes towards the learning situation: attitudes towards the language teacher and the language course.
- Motivation: effort, desire and attitude towards learning the language.

Gardner (1985) explained that motivation should be seen as goal-directed and the goal is to learn the target language. Therefore, the term orientation is used and takes the form of motivational antecedents (Gardner, 1985). Two orientations were proposed by Gardner (1985) namely; integrative and instrumental. People learning the target language in order to assimilate

with the L2 community and understand their cultural values are said to have an integrative orientation. Those who learn the language for the sake of career prospects, achieving good grades or even to make them better educated are said to have an instrumental orientation. However, the reasons for learning a language can be various and of course, these orientations are not antithetical, as a learner can have more than one orientation simultaneously and these can differ in degree over time. Drawing on the socio-educational model and the L2 motivation theory, the Attitudes Motivation Test Battery (AMTB) was developed (Gardner, 1985; Gardner & Lambert, 1972). The instrument was designed to measure L2 motivation and its associated variables which included integrative and instrumental orientation, attitudes towards language learning, parental encouragement and language anxiety. The AMTB is still today a widely known instrument that has been adopted and/or adapted by many studies on L2 motivation. Although the instrument has been modified from one study to another because of different sociocultural contexts, the main variables being investigated are likely to be the same.

Following Gardner's motivation theory, Richard Clément proposed the notion of linguistic self-confidence, one's belief in one's ability to use the L2 and accomplish one's goals, (Clément, 1980; Clément et al., 1994) which can also in a way be related to the psychological notion of self-efficacy (Bandura, 1997). The latter refers to a person's belief that they can accomplish a specific task successfully. The concept of linguistic self-confidence has been described as a "powerful mediating process in multi-ethnic settings" (Dörnyei & Ushioda, 2011, p. 43) that influences one's motivation to learn and practise the language of the L2 community. Clément et al. (1994) suggested that in a context where different language communities live together, the exposure to different languages will be the main motivational factor in learning a language from other communities, hence the rise of intercultural communication and the identification with the L2 group. Therefore, although self-confidence has a cognitive component (i.e. perceived L2 proficiency), Dörnyei and Ushioda (2011) argued

that Clément's linguistic self-confidence is mainly a socially defined construct which stands somewhat in contrast to the cognitive nature of self-efficacy.

2.7.2 Dörnyei's Framework of L2 Motivation

Although Gardner's model and motivation theory have been widely used, some limitations have been identified by researchers in the field. The theory was socially oriented and little attention was given to other aspects such as the classroom context and cognitive processes. In addition, Dörnyei (1994) criticised it for presenting motivation as a static rather than a dynamic phenomenon. Dörnyei (2005) argued that the model is not elaborate but only showed a "schematic outline" (p.68) of how motivation is connected to other individual differences and overemphasised that language achievement is heavily influenced by integrative motivation. In addition, the terminology used by Gardner to describe motivation was confusing and inconsistent, for example, integrative orientation, integrativeness, and integrative motive/motivation (Dörnyei, 2005). These criticisms gave rise to the emergence of another perspective on L2 motivation in what is known as the cognitive-situated period. The cognitive-situated approach, however, did not entirely reject the previous theories but in fact, built upon them, resulting in the development of the framework of L2 motivation (Dörnyei, 1994, 2005). Within the framework, Dörnyei further proposed two important notions: a cognitive perspective that includes language learning in relation to a learner's self-perception, abilities and values, and a micro-perspective of motivation (Dörnyei, 1994, 2005). The latter shifted from Gardner and associates' macro-perspective of L2 motivation that was focused on language learners within the L2 community to a narrower context in which it portrays "actual learning situations such as the language classrooms" (Dörnyei, 2005, p. 74). Dörnyei's framework of L2 motivation encompasses three levels:

- Language level – incorporates Gardner's notions of integrative and instrumental subsystem.
- Learner level – comprises language learning process in relation to the learner's individual characteristics which includes the need for achievement and self-confidence (anxiety, perceived L2 competence and self-efficacy).
- Learning situation level – involves situation-specific motives within the classroom settings (Dörnyei, 1994, 2005; Dörnyei & Ushioda, 2011).

As the present study investigates the effects of explicit and implicit instruction through speaking tasks, thus it is important to discuss Dörnyei's learning situation level. The situation-specific motives can be broken down into three components: course-specific, teacher-specific and group-specific motivational components.

The course-specific motivational components are mainly concerned with the syllabus teaching approach, materials used in teaching and learning tasks (Dörnyei, 1994). That said, these components connect teachers and the students. In other words, if the learning process is like a 'tunnel', with teachers at one end and students at the other, then these components (i.e. approach, materials etc.) are what pass through the tunnel from one end to another. In addition, these components are mostly determined by the teachers because the course, approaches and the materials used in the classroom are tools that help teachers to exert positive motivational impact on students. Dörnyei (1994) described four main course-specific motives which were originally proposed by Crookes and Schmidt (1991): interest, relevance, expectancy and satisfaction. In short, interest concerns intrinsic motivation and sustaining learners' curiosity and desire in learning; relevance refers to the extent to which learners' personal needs, values and/or goals can be met in the course; expectancy is related to learners' overall expectancy for success associated with self-confidence and self-efficacy at a general level; satisfaction includes extrinsic rewards such as enjoyment. From the teacher's point of view, these four

components are considered helpful in relation to the present study as it concerns not only the design of motivating speaking tasks but also their implementation using two different forms of language teaching.

Evidently the teacher-specific components involve the role of the teacher, who is seen as a vital character in respect of learners' learning outcomes in classroom settings. Three teacher-specific aspects are identified in Dörnyei's framework. The first is affiliative drive which refers to students trying to do well in order to impress the teacher, for example, if the teacher is likeable and appreciated by the students, thus there is a tendency for the students to perform. The second is authority type that is concerned with the teacher's approach with the students whether it be autonomy-supporting or controlling. Therefore, students' learning motivation might either be positively or negatively influenced by the teacher's style of teaching. The final aspect of this component is the socialisation of motivation that is, whether the teacher continuously stimulates students' interest and motivation by being a role-model, raising their metacognitive awareness and giving constructive feedback. Although the present study does not focus on the teacher as the participant, the two different types of instruction implemented in the study could influence the students' perceptions of the role of the teacher in the classroom and hence affect their overall motivation.

Finally, group-specific components are related to group dynamics, as, according to Dörnyei (1994), "classroom learning takes place within groups as organisational units" (p. 278). The components could be briefly summarised as how members of the group work together in achieving the goal that matches the agreed norms within a group; different goal structures (e.g. competitive, cooperative and individualistic) may have different effects in encouraging leaners' motivation. In addition, the component also implies that reward systems have a strong influence on individual learners' motivation. For the present study, the concept of 'group' might differ from one speaking tasks to another. As the intervention involved a diverse range

of themes and followed the course outline which incorporates specific learning objectives, thus students might not be working in groups at all times. The intervention combined individual work, pair and group work while doing different speaking tasks. Therefore, it could be assumed that students' motivation would be affected in various ways according to the task.

Dörnyei's multifaceted framework of L2 motivation stimulated further studies of L2 motivation which incorporated both individual learners and classroom-specific components. As the present study investigated the effects of two different language teaching, hence it involved students' attitude and motivation towards the learning situation. In other words, these two different language teaching (i.e. TBLT and TSLT) can be categorised as variables in the learning situation motives which could affect students' motivation to learn and speak English. Therefore, this framework can be applied to classroom-based language learning situations and can provide valuable understanding into students' motivation in the present study.

2.7.3 Bandura's Self-Efficacy Theory

It is within human nature to desire a sense of agency, a belief that one has the autonomy and power to exert control over one's life (Bandura, 1989, 1997, 1999). Agency embodies the concept of empowerment and self-efficacy is seen as an essential mean to become empowered (Schunk & DiBenedetto, 2016). Self-efficacy, which is commonly drawn on in L2 motivation models (Dörnyei, 1994), is an important variable of human agency in the social cognitive framework developed by Bandura, affecting one's decision to act and to persevere in performing tasks (Bandura, 1997). Self-efficacy denotes the "belief in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3), hence is concerned with expectations for success. Bandura (1997) distinguished between two types of expectations for success as shown in Figure 2.6 below, which he termed as efficacy beliefs and outcome expectancies.

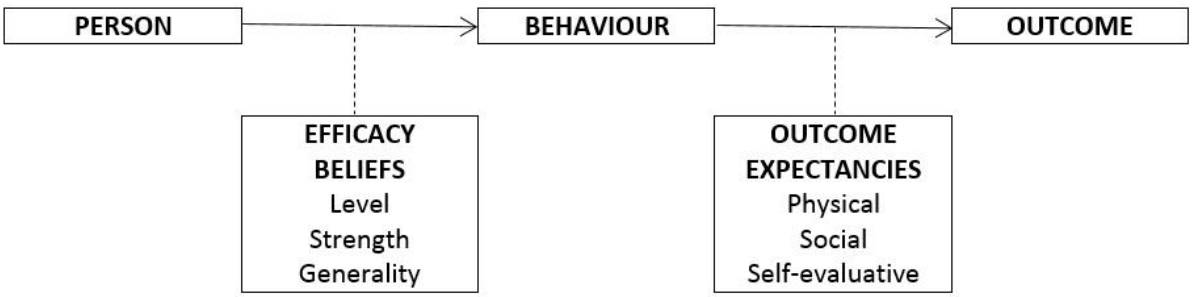

Figure 2. 6: Relationships between efficacy beliefs and outcome expectancies (Bandura, 1997, p. 22)

In this perspective, outcome is the result of one's behaviours. Similarly, the anticipated outcome is highly dependent on how individuals evaluate their own capability in performing the given task (Bandura, 1997). Self-efficacy can affect a person's choice whether they either participate in a task for which they believe they are efficacious or avoid it entirely (Schunk & DiBenedetto, 2016). In addition, self-efficacy also determines how much effort a person is willing to put in during task performance, how long they will persist and how resilient they are when encountering challenges (Schunk & DiBenedetto, 2016). To put these ideas into an educational context, learners with a low sense of self-efficacy beliefs see the task as a 'threat' and dwell on their own weaknesses which could lead them to believe themselves likely to fail the given task. Learners with a high sense of self-efficacy beliefs, on the other hand, believe that they are capable of performing the given task and sustaining their effort even if they may face difficulties (Dörnyei & Ushioda, 2011).

Bandura (1997) underlined in the social cognitive framework the triadic interaction between person (cognitive and affective), biological events (behaviour) and the environment. The relationship between person and behaviour is heightened with the presence of self-efficacy which can be stimulated by the environment. That said, self-efficacy emerges from four main sources, outlined in Bandura (1997) as: enactive mastery experiences, vicarious experiences,

verbal persuasion, and physiological and affective states. These sources could simply be summarised as:

- Enactive mastery experiences – the level of efficacy depends on having prior successful experiences. If a person had previously performed a task well, s/he is more likely to feel competent in the upcoming task.
- Vicarious experiences – occur when the person believes they have the ability to perform the task after observing other similar people's successful performance.
- Verbal persuasion – refers to the influence of appraisal and encouragement from others. A person who received appraisal and/or constructive feedback is likely to have a high level of self-efficacy
- Physiological and affective states – relate to the person's emotional states. Emotional reaction to a certain task (e.g. excitement, anger, anxiety) can lead to either a positive or negative judgement of one's ability to complete the task.

In the present study, self-efficacy refers to the students' perceived efficacy of their English language knowledge and speaking skills. The aim which is being measured against, therefore, is not whether students understand the speaking task, but rather do they have the ability to speak, perform in speaking tasks and interact with others.

2.8 Willingness to Communicate (WTC)

Many language teachers may have encountered challenges in encouraging students to practice the target language, especially speaking. From the learners' perspective, they may experience learning difficulties such as speaking anxiety which could cause them to be unwilling to speak and participate in classroom activities and hence avoid entirely performing in tasks. In some cases, students tend to avoid speaking in communicative tasks even though they have a high level of communicative competence. This could indicate that there might be a "layer of

mediating factors between having the competence to communicate and putting this competence into practice" (Dörnyei, 2005, p. 207). As previously mentioned, research into individual differences has demonstrated the influence of affective constructs on learning outcomes. A recent addition to the range of affective variables explored is willingness to communicate (WTC) (Clément, Baker, & MacIntyre, 2003; MacIntyre, Dörnyei, Clément, & Noels, 1998; McCroskey & Baer, 1985; McCroskey & Richmond, 1990). WTC involves variability in talking behaviour which according to McCroskey and Richmond (1990) is:

> *Some people talk very little, they tend to speak only when spoken to-and sometimes not even then. Others tend to verbalize almost constantly. Many people talk more in some contexts than in others, and most people talk more to some receivers than they do to others* (p. 72).

The WTC construct was introduced by McCroskey and associates to address communication in the L1 context (McCroskey & Baer, 1985; McCroskey & Richmond, 1990) which was conceptualised within a personality traits orientation. It began to receive increasing interest in L2 research as it was discovered that the level of WTC among L2 learners affects the quality and quantity of their L2 speaking outcomes (Clément et al., 2003; Yashima, 2002) which can in turn, assist successful L2 acquisition. Nevertheless, the WTC construct received a lot of criticisms as it focused only on personality variables, thus situation-based variables were added in the WTC construct and were adapted in the L2 context (Clément et al., 2003; MacIntyre et al., 1998; Yashima, 2002; Yashima, Zenuk-Nishide, & Shimizu, 2004). The situation in the L2 context is considered more complex than the L1 because the level of L2 proficiency particularly one's L2 communicative competence is seen as an additional modifying variable (Dörnyei, 2005). Therefore, MacIntyre et al. (1998) argued that WTC in L2 needs to be conceptualised as a combination of both personality traits and situation-based characteristics, which are

defined as one's "readiness to enter into discourse at a particular time with a specific person or persons, using a L2" (p. 547). Accordingly, MacIntyre et al. (1998) proposed a multi-layered pyramid called the heuristic model based on the belief that "authentic communication in L2 can be seen as the result of a complex situation of interrelated variables" (p. 547). The model illustrated in Figure 2.7 consists of a various range of communicative, social psychological and linguistics variables that were categorised into six main layers.

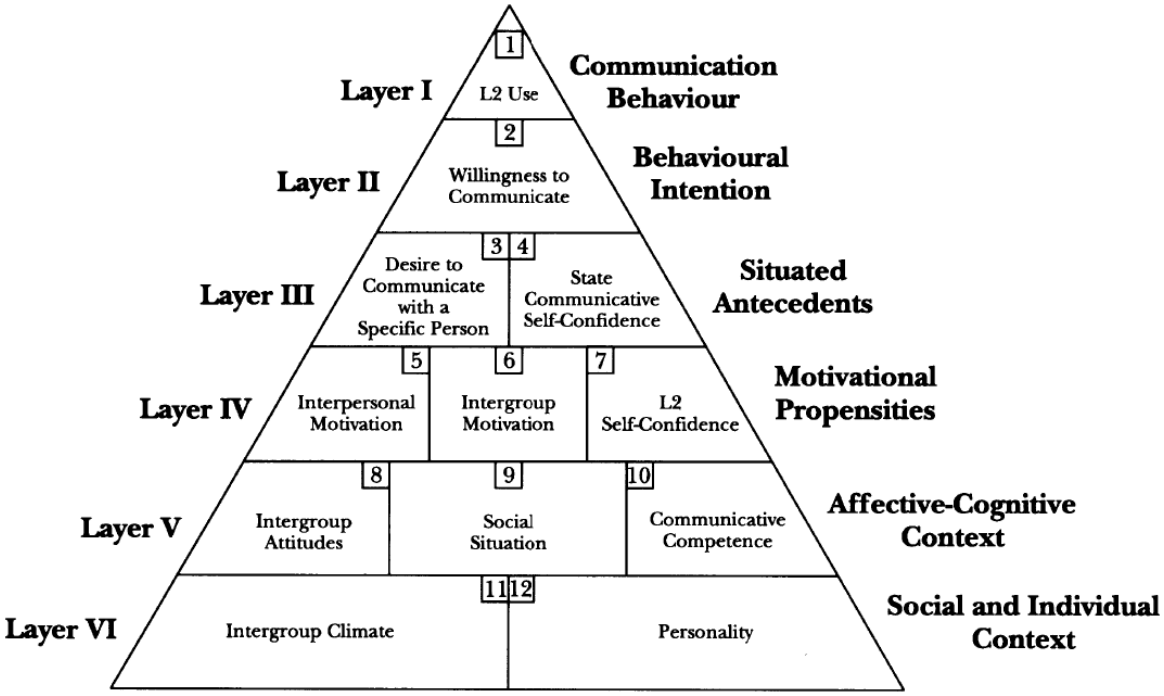

Figure 2. 7: Model of variables influencing WTC (MacIntyre et al., 1998, p. 547)

The first three layers (I, II and III) are situational and changeable variables that influence a person's WTC, whereas the following three layers (IV, V and VI) signify more enduring variables influencing a person's WTC. The model could be summarised as follows:

- Layer I: Communication Behaviour – L2 use is placed on top to indicate the ultimate goal of the learning process is for learners to produce the L2 language.

- Layer II: Behavioural Intention – concerns learners' readiness to engage themselves in discourse which suggests that learners with a high level of WTC are more likely to use the L2 in any given situation.
- Situated Antecedents – comprises two situational antecedents: desire to communicate with a specific person (e.g. physically nearby, frequently contact, attractive) and the state of communicative self-confidence (e.g. perceived communicative competence and the lack of communicational anxiety).
- Motivational Propensities – involves interpersonal motivation (e.g. controlling or dominating versus affiliating with the communicative discourse), intergroup motivation (i.e. behaviour towards members of a particular group) and self-confidence (e.g. perceived L2 competence and L2 anxiety).
- Affective-Cognitive Context – consists of intergroup attitudes (e.g. integrativeness, fear of assimilation and motivation to learn the L2), social situation (e.g. age group, gender, setting, purpose of communication and topic of discussion) and communicative competence (e.g. linguistic, discourse, sociolinguistic and strategic competence) which are heavily influenced by the L2 community.
- Social and Individual Context – includes intergroup climate (e.g. structural characteristics and perceptual and affective correlates) and personality (e.g. extraversion, agreeableness, conscientiousness, emotional stability, and openness to new experiences).

Therefore, the intricacy of the relationship between variables and the learner indicate that learners' communicative experience in a certain situation may not automatically be transferable to another situation (MacIntyre et al., 1998). Therefore, a person's level of WTC differs

depending on the social situation in which the event takes place and to whom the person is engaged with.

2.9 Studies in L2 Motivation and WTC

Integrativeness versus Instrumentality

Several studies have adopted Gardner's motivation theory to investigate the level of integrativeness and instrumentality in the L2 context, particularly in a multiracial country like Malaysia (Hong & Ganapathy, 2017; Muftah & Rafik-Galea, 2015). For example, Hong and Ganapathy (2017) studied whether instrumental or integrative motives play an important role in promoting ESL secondary school students towards English language learning. A total of 12 secondary school students were recruited for a focus group interview and formed two groups of six students. The interview questions were based on the AMTB instrument that focused on integrativeness, instrumental motives and also students' perception of language learning difficulties on listening, speaking, reading and writing. Results showed that instrumental motivation was a more important factor for students' English language learning. The study also suggested that even though students were instrumentally motivated, they did not put much effort into learning the English language such as doing their homework when they were asked to. However, students who were interactively motivated were more eager in reading beyond their textbook and learning English outside the classroom. In addition, the study revealed that difficulties in English vocabulary and grammar were the main cause of students feeling demotivated in learning English.

These findings echo those by Muftah and Rafik-Galea (2015) who used a survey with close and open-ended items, similarly adapted from the AMTB, with 182 pre-university non-English students. The quantitative results showed that students were both highly instrumentally and integratively motivated with the former being slightly more important. Through the qualitative

data, the study also suggested that instrumentality is more prominent in situations where there is little desire to integrate with the target language community and students are trying to preserve their identity. In addition, the findings also discovered that students felt demotivated when they had few opportunities to practise English outside the classroom and when they had to spend a lot of time and effort to learn the grammar of the language. Therefore, this might suggest that students in the context of Malaysia are more instrumentally motivated, and that their level of motivation is higher when they have specific goals for learning such as to get good grades rather than having general goals for learning the L2.

Learning Situation Motives

As the study focuses on two different language teaching namely TBLT and TSLT, thus teacher or teaching-related variables were discussed. These variables, were mainly the role of the teacher, teaching approach and the type of instruction, which play a big role in students' L2 motivation (de Smedt, Graham, & van Keer, 2018; NamazianDost, Bohloulzadeh, & Pazhakh, 2017). A study by de Smedt et al. (2018) used an experimental, pre-/post-test design with 206 fifth-and sixth-grade students to investigate the impact of explicit instruction (EI) and peer-assisted (PA) writing on students' writing motivation and self-efficacy for writing. These students were divided into five groups including control, explicit instruction + writing individually (EI+IND), explicit instruction + peer-assisted writing (EI+PA), peer-assisted writing (PA) and individual writing (IND) only. Each group was given writing lessons with different conditions that lasted for five weeks. The results revealed that groups that received EI and EI+PA in particular had higher motivation compared to those groups which did not, followed by the PA group. However, although groups that received EI had higher writing motivation, the study suggested that they had controlled writing motivation. In other words, how high their motivation was depended on how much input they received from their teacher,

hence they had a sense of dependency. In contrast, the PA writing group encouraged students' innate needs, hence creating autonomous writing motivation.

A study by NamazianDost et al. (2017) obtained different results. The study investigated the effect of TBLT on the motivation and grammatical achievement of 80 EFL junior high school students. Students were assigned to two groups: experimental (grammar through TBLT) and control (conventional grammar teaching by drilling), in which the sessions lasted for 12 weeks. A grammar test and motivation questionnaire were administered before and after the intervention. The findings showed that students in the experimental group which did not receive explicit instruction were highly motivated and significantly improved in their grammar compared to the control group. In other words, instruction used in the classroom can affect students' motivation. The study suggested that the TBLT approach encouraged the students in the experimental group to be more creative which was shown by a variety of grammar usage in their post-test; this stimulated students' motivation and willingness to learn the language better. Therefore, different approaches and instruction types may affect students' overall motivation differently which leads students to use the language differently.

Self-Efficacy

In the Malaysia context, self-efficacy levels towards English language learning have been investigated and compared between first year and final year undergraduate students (Idrus & Salleh, 2008; Mohamed Khatib & Maarof, 2015). Through a questionnaire, Mohamed Khatib and Maarof (2015) investigated self-efficacy in speaking English among 60 technical polytechnic students based on two different semesters (first and fifth semester). The questionnaire assessed perceived ability, attitude and aspiration. The results showed that students in their final semester showed a significantly higher level of self-efficacy than the first semester students. This might be due to the fact that students with high self-efficacy level had gone through industrial training or internship during their first semester, hence through

vicarious experiences they were exposed to real world communication and training in the workplace such as conversations with experienced colleagues and presentations in department meetings. However, students in their first semester were found to have low self-efficacy as they lacked exposure to real-life situations, in so far as they had just started their tertiary level and were not exposed to situations using various types of communication skills.

However, Idrus and Salleh (2008) who studied 338 engineering/technical students in a private university self-efficacy, discovered that both first and final year students showed a high level of self-efficacy although as in Mohamed Khatib and Maarof (2015) the latter group scored more highly than the former. The study suggested that the first year students' self-efficacy came from successful experiences prior to university, as they had just recently graduated from pre-university level with distinction for the English subject and during the selection process of university entrance, they passed the interview which was conducted in English. As for the final year students, their self-efficacy in their speaking ability was higher as they had had longer exposure to the language during their university years. In addition, they had taken all courses in English with various speaking tasks and had gone through an internship which further heightened their level of self-efficacy. Therefore, students' positive experiences and opportunities to observe other successful and experienced people had a positive impact on their self-efficacy.

Willingness to Communicate

A study by Abdul Razak, Nimehchisalem, and Abdullah (2018) reported that high levels of ethnic group affiliation (i.e. how closely a person feels attached to their native culture and language) resulted in low levels of WTC in English among undergraduates students. Through questionnaire, data from 39 students from three ethnic groups: Malay, Chinese and Indian were obtained. The study suggested ethnic groups often pressurised their members to behave in a certain way in order to protect their existence, native language and identity. Therefore, students

felt a responsibility to preserve their identity which led them unwillingly to communicate in another language.

Besides the involvement of ethnicity, S.-J. Kang (2005) stated that a person's situational WTC in L2 can dynamically emerge and fluctuate during a conversation situation. In that study, four Korean students, who were non-native speakers of English, were selected and they were exchange students for an English programme in the United States. Through qualitative methods (interview, stimulated recalls and observation), it was found that students felt excited to communicate when they were given a topic related to personal experience (e.g. family, culture) and if the other interlocutors helped them to improve their English. In addition, a sense of responsibility to communicate emerged when students had to clarify misunderstandings or wrong information during a conversation and if the topic was one that students perceived to be useful and important. However, the study suggested that students' level of WTC decreased when the situation involved many interlocutors, speaking with a non-native speaker who was much fluent than they were and when they lack of background knowledge of the topic.

Similarly, Philp and Cao (2006) investigated trait-like and situational WTC on eight university students of which were non-native speakers of English from various background. The data were collected through interview, questionnaire and classroom observation. The results showed students' unwillingness to communicate was caused by lack of self-confidence resulting in reticence in class. Other factors that would influence students' WTC behaviour which were in line with S.-J. Kang (2005) included group size, familiarity with the interlocutors, familiarity of the topic and interest towards the topic.

In summary, the seven studies related to motivation and willingness to communicate offer several findings. Studies have shown that students in the ESL context were instrumentally motivated (Hong & Ganapathy, 2017; Muftah & Rafik-Galea, 2015). This was because

students in both studies of Hong and Ganapathy (2017) and Muftah and Rafik-Galea (2015) had the notion that learning the English language was to get a good grade in their studies or pass the subject. In addition, studies discovered that students in their final year who had vicarious experiences through industrial training showed high levels of self-efficacy compared to first-year students (Idrus & Salleh, 2008; Mohamed Khatib & Maarof, 2015). Findings of Mohamed Khatib and Maarof (2015) and Idrus and Salleh (2008) showed that final year students scored a high level of self-efficacy due to the vicarious experience they had during industrial training compared to first-year students who were not exposed to the real working world. In addition, studies revealed that students' willingness to communicate is not only influenced by personality traits (Philp & Cao, 2006) but also situational-based variables (Abdul Razak et al., 2018; S.-J. Kang, 2005). The main reason students were not willing to communicate with others is due to lack of self-confidence (Philp & Cao, 2006) and to preserve their identity (Abdul Razak et al., 2018). The willingness to communicate depended on three main factors which were excitement (e.g. topic of discussion, interlocutors); responsibility (i.e. the obligation to deliver and understand a specific message) and security (i.e. feeling safe) (S.-J. Kang, 2005). However, these studies were mainly an exploratory in nature where methods such as interview and questionnaire were employed rather than an experimental study, and did not involve a comparison between pre- and post-test. Therefore, the distinction between pre-existing motivation and motivation produced from the instruction cannot be made.

Furthermore, findings of previous studies have explored the learning situation motive mainly from the role of the teacher, instruction and task employed in the classroom which had impacted students' L2 motivation (de Smedt et al., 2018; NamazianDost et al., 2017). The findings of de Smedt et al. (2018) showed that students that received explicit instruction in writing were more motivated from pre- to post-test but their motivation was controlled and dependent on the teacher, whereas students who peer-assisted and did not receive any explicit instruction were

more autonomously motivated. NamazianDost et al. (2017) found that learning grammar through TBLT had a positive effect on students 'motivation and grammar usage over time. Although the previous studies provided a robust investigation using a quasi-experimental pre-/post-test design, the studies were mainly focused on the effects of writing and grammar task on students' motivation towards learning writing and grammar. Very few have considered motivation in relation to speaking tasks. Therefore, the aim of this study is to investigate the extent to which TSLT and TBLT have an impact on students' motivation towards learning and speaking English in terms of attitude, integrativeness, instrumentality, linguistic self-confidence and learning situation.

2.10 Research Questions

In short, existing research in relation to the main variables of the study showed some limitations such as investigating one or two dimensions of CAF only (de la Fuente, 2006; S. Li et al., 2016; Van de Guchte et al., 2017), using only global measures of CAF (Ellis et al., 2019; S. Li et al., 2016), and focusing on specific grammatical and vocabulary items (Ahmadian, 2012; de la Fuente, 2006; Ellis et al., 2019). As for motivation, some studies did not employ a quasi-experimental design (Abdul Razak et al., 2018; Hong & Ganapathy, 2017; Mohamed Khatib & Maarof, 2015) and only focused on writing and grammar tasks rather than on speaking and on the impact of different teaching approaches (de Smedt et al., 2018; NamazianDost et al., 2017). Therefore, there is a need to address these issues. In the light of the literature reviewed and discussed above, the present study proposes to address this gap by considering the following research questions:

1. What are the effects of the implementation of two different forms of language teaching (TBLT and TSLT) on:
    a. the development of students' English speaking performance over time?

b. students' motivation towards learning and speaking English over time?

2.11 Summary

This chapter discussed the literature relating to TBLT, speaking and the production of speech, the role of explicit and implicit instruction in SLA and L2 motivation. First, TBLT was discussed as it is one of the approaches used in Malaysia, the context of the study, and has been contrasted with TSLT that implements explicit instruction within tasks. Second, the main variable of the study which is speaking was described in relation to L1 and L2 speech production models and their dimensions which include complexity, accuracy and fluency. Next, the role of explicit instruction was also discussed as it is one of the main characteristics of TSLT in comparison with implicit instruction. Studies related to TBLT and TSLT were also discussed. Finally, the chapter reviewed theories in L2 motivation towards language learning and willingness to communicate, and their significant role in speaking. All of these sections were followed by discussion of previous studies addressing the variables.

CHAPTER THREE: METHODOLOGY AND METHODS

3.1 Introduction

This chapter discusses the research paradigm and how it influences the choices of instruments used in this study. The chapter also presents the research design, methodology and methods to address the research questions of the study. In addition, participants of the study together with the intervention, data collection and the ethical consideration are also discussed.

3.2 Research Paradigm

The term research paradigm can be defined in various ways, such as worldview (Creswell, 2013b), research methodology (Neuman, 2006) or epistemologies and ontologies (Crotty, 1998). In this thesis, a paradigm is taken as a system that guides the way a researcher does things and establishes a set of practices. Therefore, understanding the selected paradigm facilitates the choice of precise methods of data collection, the procedures for analysing and interpreting the collected data and selecting a sample for the study. In the social sciences, it is sometimes argued that there are two main paradigms, namely qualitative and quantitative, but this can be viewed as an oversimplification as both primarily emphasise "data rather than foundational beliefs and assumptions" (J. W. Willis, 2007, p. 8). Taking a more nuanced position, even though different terms may be used from scholar to scholar, three paradigms are prominently identified, namely post-positivism, critical theory and interpretivism (Greene, Benjamin, & Goodyear, 2001; Guba, 1990; Smith, 1989). Quantitative research may fall into the paradigm of post-positivism, whereas qualitative research would belong in both critical and interpretivism. However, for mixed methods research, Teddlie and Tashakkori (2009) suggested a pragmatism paradigm which combines both quantitative and qualitative research.

Pragmatism focuses on finding a solution to every research problem using all the most appropriate approaches available (Creswell, 2013b). In other words, pragmatism uses pluralistic approaches to derive knowledge about a problem (Cohen, Manion, & Morrison, 2018; Creswell, 2013b) and this applies to the mixed methods approach where researchers draw assumptions from both quantitative and qualitative types of research. Mixed methods research addresses both the 'what' and 'how' or 'why' types of research questions which is essential if the intention of the research is to understand the different explanations of outcomes in various ways (Cohen et al., 2018). Therefore, pragmatism fully opens the door to multiple methods, different perspectives and various assumptions, as well as various data collection methods and analysis. Importantly, pragmatists do not see the world in only one way by restricting themselves to one method alone. By selecting more than one method, they achieve full autonomy of what is being researched, and how the procedure is followed and what technique is used in order to achieve their research purposes (Creswell, 2013b; Gray, 2014).

Both quantitative and qualitative research have their advantages and limitations. In short, quantitative allows data to be collected on a bigger scale and statistical analysis is considered a more reliable way to observe changes, but it may lack validity in so far as that it may not offer the opportunity for clarification or in-depth exploration (Queiros, Faria, & Almeida, 2017). As for qualitative research, it provides rich and in-depth information from the participants under investigation but the data collected can be ambiguous (a feature inherent in human language) and would not always be generalizable due to the small sample sizes often used (Ochieng, 2009). As for the present study, both quantitative and qualitative methods were used to obtain a more robust analysis because "neither quantitative nor qualitative methods are sufficient, by themselves, to capture the trends and details of a situation" (Ivankova, Creswell, & Stick, 2006, p. 3). Consequently, the present study adopts a sequential explanatory design which consists of two phases shown in Figure 3.1 below. The first phase of the study starts

with quantitative data collection and analysis and is followed by the subsequent collection and analysis of qualitative data (Creswell & Clark, 2007). Greater emphasis is given to the quantitative methods (i.e. in Figure 3.1 QUAN is capitalised), hence the qualitative methods (i.e. qual is lower case) and data are supplementary and help to explain and gain a deeper understanding of the quantitative findings (Creswell & Clark, 2007).

Figure 3. 1: Mixed methods sequential explanatory design (Creswell & Clark, 2007, p. 73)

The study investigated the effects of two different forms of language teaching (TBLT and TSLT) on students' speaking performance and motivation. Quantitative methods such as a speaking test, questionnaire and statistical analysis were used to investigate whether the intervention could improve students' speaking performance and motivation over time. Meanwhile, the qualitative methods, including the interviews and the qualitative analysis of the speaking task data (linguistic and thematic analyses of participants' speech production) were used to gain a thorough and comprehensive view of students' speaking outcomes and their motivation towards learning and speaking English. In the present study, mixed methods were employed together with the quasi-experimental pre-/post-test design. The following sections will provide details of the quasi-experimental pre-/post-test research design which was adopted as the main design of the study.

3.3 Quasi-Experimental Pre/Post-test Research Design

An experimental research design is usually associated with the use of quantitative methods, but qualitative data can also be collected in a quasi-experiment (Bryman, 2012). In order to address the research questions, the present study employed a quasi-experimental pre-/post-test research

design. An advantage of using a quasi-experimental design is that it has high ecological validity in which it does not manipulate or interfere with the classroom context by placing participants in groups that differ from those in which they normally work (Mackey & Gass, 2005). To reiterate, the aims of this study were to investigate the effects of two forms of language teaching (TBLT and TSLT) on students' speaking performance and motivation towards learning and speaking English. Two intact groups were selected in this study as it was not possible to randomly assign participants to certain groups (Gay, Mills, & Airasian, 2011). The two groups were randomly assigned to two different experimental groups: the EIST (explicit instruction + speaking tasks, i.e. which received TSLT) group (N = 30) and the ST (speaking tasks only, i.e. which received TBLT) group (N = 29) (see Section 3.5.2 for the differences between the instruction used in both groups). The participants of the study were first-year undergraduate ESL students in a technical/engineering university in Malaysia (see section 3.4 for details of the participants in the study).

The quasi-experimental study consisted of three main phases: pre-test, intervention and post-test as presented in Figure 3.2 below. The pre-test was administered two weeks prior to the intervention. Within the pre-test phase, four instruments were used to measure the key dependant variables. The vocabulary test only functioned as a covariate in the statistical analysis while the questionnaire was used to investigate students' L2 motivation towards learning and speaking English. These two instruments were administered to the students at the first meeting in the first week followed by the speaking test and the interview. Two weeks were required for the speaking test and interview as they were conducted individually and also students were not from the same undergraduate engineering/science majors, hence they had different scheduling times. The pre-test instruments were administered before the intervention to seek information about students' past motivation towards learning English and speaking English, and to measure students' speaking performance level at that time before the

implementation of the intervention. The post-test instruments were used to gather information if the intervention had any effects on students' speaking performance and motivation, thus data collection took place immediately after the conclusion of the intervention to avoid the influence of any future potential variables and to minimise the effect of the time gap.

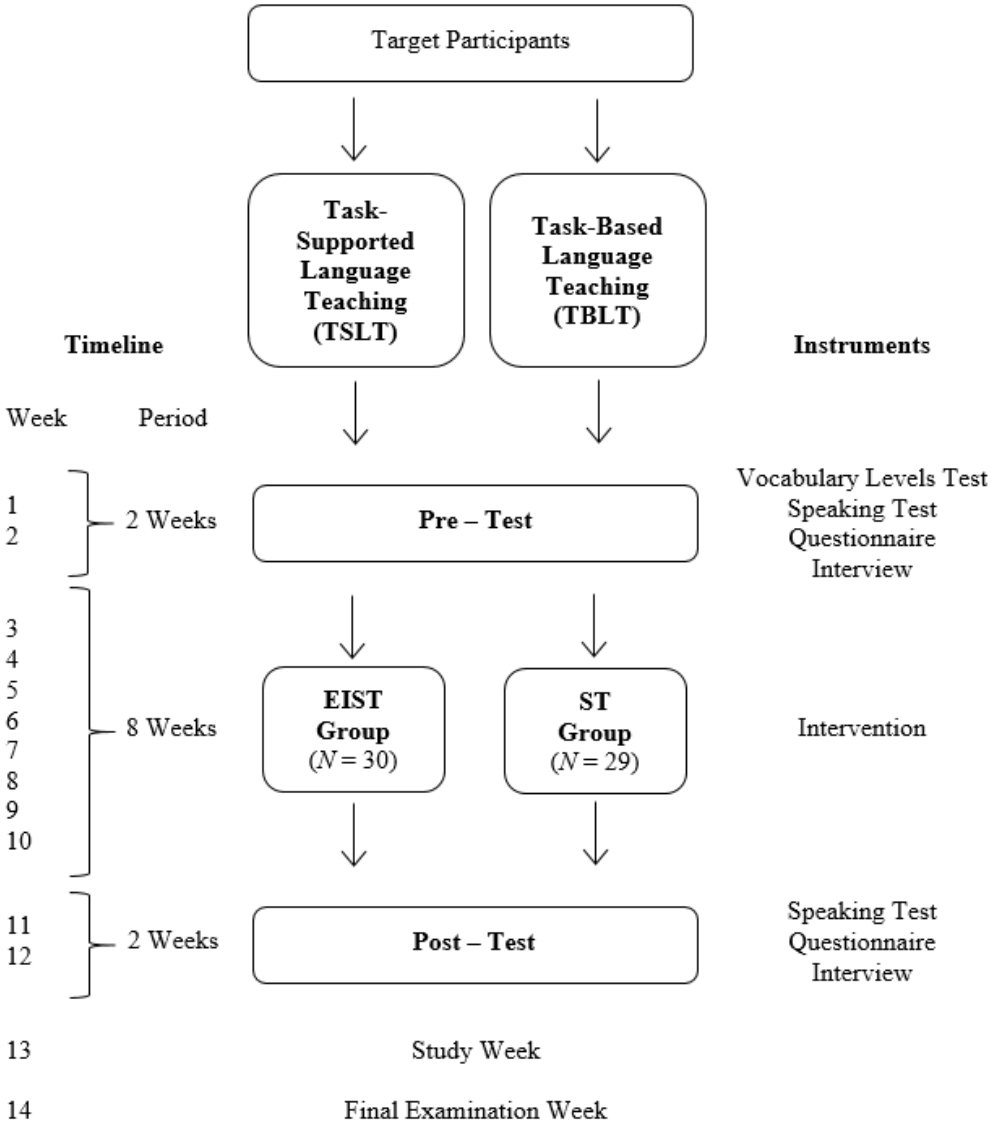

Figure 3. 2: Design of the present study

The study was conducted within a remedial English course, which is Fundamentals of English language, (henceforth, FoEL) taught by the researcher. The course is structured according to the MUET that covers all four skills of English (Sulan et al., 2014). Initially, the course is a two-hour class, conducted twice a week for a 12-week period. However, speaking skills could

not be taught entirely throughout the whole semester as priority had to be given to other skills. Therefore, the speaking intervention included two-hour speaking lessons once a week over an eight-week period (i.e. two hours per week). Although the intervention was conducted for only two hours a week for eight weeks, given the time constraint of every student, pre/post-test data collection, and to teach other skills beside speaking, as well as the mid-term break (one week) and study week (one week before final examination), it was thought to be realistic and suitable for the present study. Previous studies have implemented even shorter interventions on just one day for two hours (Ellis et al., 2019; S. Li et al., 2016) or 50 minutes (de la Fuente, 2006).

3.4 Participants of the Study

The target participants (N = 59) were first-year undergraduate engineering/science students, aged 19 to 22 years from a public technical/engineering university located in the east coast of Malaysia. Since it is a technical/engineering university, students in this study were from various backgrounds of engineering and sciences (e.g., computer science, biotechnology, civil, chemical, mechanical, electrical engineering etc.) programmes. Although students were in different majors, they were considered to be the appropriate participants for the study, because:

- Students had received approximately 11 to 13 years of ESL education (six years in primary school, five years in secondary school, and two years in the sixth form or one year in matriculation before entering university). Students' exposure to the English language enabled them to practice and comprehend the language well. Therefore, students had sufficiently appropriate level for the nature of the instruction implemented in this study.
- Students had experienced and practised speaking which is one of the skills tested in the Malaysian University English Test (MUET). This because the speaking tasks in the intervention and the speaking tasks were designed based on the MUET speaking test format.

- Unlike second to fourth-year students, first-year undergraduate students were selected as they encountered fewer pressures from taking heavy-content courses such ESP courses.

The selection of participants was conducted through convenience sampling as the main criterion of selecting the sample related to accessibility (Lavrakas, 2008). The language centre of the university was chosen in which to conduct the study because the researcher had previously worked as a language teacher there, hence it was convenient to gain access to contact the key person in charge (i.e. dean of the centre and head of English department). Two intact groups were assigned by the head of the English department, and these groups were enrolled in the FoEL course. The FoEL course is a compulsory-to-pass course for first-year undergraduate students who did not meet the MUET requirement (i.e. who had achieved Band 3 and below). As the researcher had no control over the students' enrolment in the FoEL course, it was decided to retain the original intact groups rather than allocating students into specific groups based on their English proficiency level (e.g., MUET score - band 1 to 3). It is important to note that courses in the university are not fixed, and students have the option to choose their courses according to the credit hour and schedule, hence the tendency of having diverse students from other engineering/science majors in a class and a mixed level of English language proficiency (i.e. Band 2 and 3). Therefore, it was concluded that assigning them to different groups would disrupt students' learning to move them from their normal learning environment and would interfere with the ecological validity of the study.

As there were only two groups available for the study, these groups were randomly assigned to two experimental groups: EIST (explicit instruction + speaking task) group and ST group (speaking task only) group. Originally, both groups comprised 30 students each. However, one student in the ST group did not participate until the end of the study and thus the researcher had to disregard the student data from the study. Consequently, there were 30 students for the

EIST group and 29 students for the ST group. For a clear representation, the demographic of all participants of the study is presented in Figure 3.3 below.

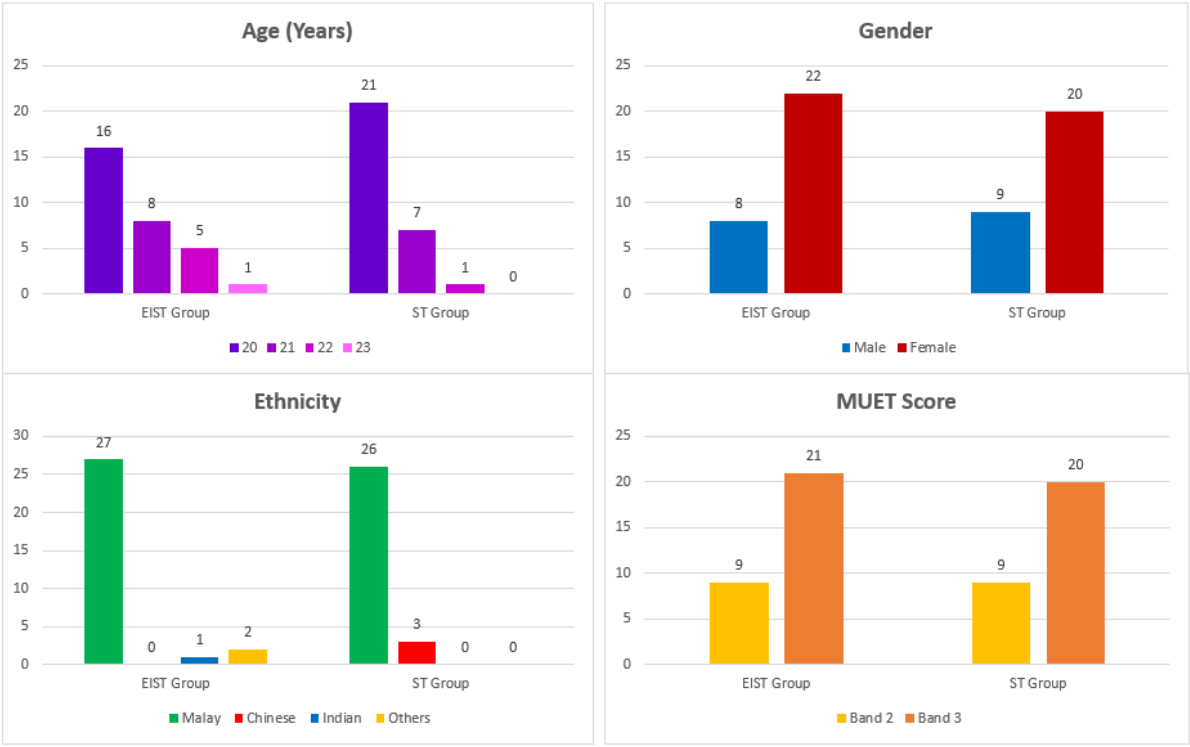

Figure 3. 3: Demographic details of the participants in the study

Both groups were taught by the researcher himself as he had experience in teaching the FoEL course for nearly two years. In addition, this is to avoid other variables such as teacher variables (e.g. gender, age, ethnicity, personality, teaching experience etc.) and to ensure the fidelity to experimental conditions.

3.5 Speaking Intervention of the Main Study

In developing the intervention, the focus was on designing the speaking tasks based on several criteria which included identifying the learning objectives and outcomes, selecting the types of tasks and engaging appropriate teaching materials (Reinbold, 2013) in order for all the lessons to run efficiently. The intervention was carefully developed by reviewing literature on explicit

instruction and task-based activities, and adopting the syllabus from the FoEL course curriculum, as discussed in the following section.

3.5.1 Reviewed Course Curriculum

The intervention in this study followed the FoEL course curriculum specification (Sulan et al., 2014). The researcher had not changed the syllabus (e.g. topics, themes, content, grammar items etc.) as it was a condition set by the department to implement the original syllabus. Even though the syllabus remained unchanged, the researcher was given full autonomy to use any kind of materials, approaches and activities to teach the course syllabus accordingly. According to Nation and Macalister (2010), adaption is needed when the current course book does not include principles, activities or materials that the teacher feels should be applied. In other words, adaptation is to alter some of the internal characteristics of the course book to meet the appropriacy of the teaching materials in context (McDonough, Shaw, & Masuhara, 2013; McGrath, 2016). Since the present study employed two forms of language teaching (TSLT and TBLT), the materials, approaches and tasks from the original course book were adapted to suit the purpose of the study without interfering with the content of the current syllabus. The syllabus for the FoEL (Sulan et al., 2014) as shown in Table 3.1 below was adopted in the present study.

Table 3. 1 *: Outline of the FoEL syllabus (Sulan et al., 2014)*

Lesson / Week	Content	
	Themes	**Grammar Items**
1	Environment	Nouns and adjectives
2	Sports	Verbs and adverbs
3	Education	Pronouns (personal, reflexive, possessive)
4	People and lifestyles	Article and determiners
5	Leisure and pastimes	Simple present and past tense
6	Tourism and travel	Present and past continuous tense Present and past perfect tense
7	Health	Modal verbs
8	Science and technology	Future tense

3.5.2 Instruction

The lessons conducted in the intervention were divided into three main stages: pre-task, during-task and post-task which reflects the chronology of the TBLT lesson (Ellis, 2009b). Similar to the PPP stages, these three stages have their own purposes. According to Ellis (2009b), the pre-task stage introduces the topic prior to the actual task which includes explicit instruction and it also serves to attract students' interest towards the lesson. During-task includes practice under time pressure and post-task entails follow-up after the task performance and encourages reflection and feedback (Ellis, 2009b).

3.5.2.1 Instruction for the EIST Group

As for the EIST group, explicit instruction and grammar/vocabulary exercises were given to the students prior to the speaking tasks (i.e. pre-task stage). All of the explicit information relating to grammar and vocabulary were presented using PowerPoint slides followed by a short grammar/vocabulary exercise in the form of a worksheet. The explicit instruction took approximately 10 minutes which was considered appropriate as previous studies have been conducted within the range of 10 to 15 minutes for explicit instruction (de la Fuente, 2006; Ellis et al., 2019; S. Li et al., 2016).

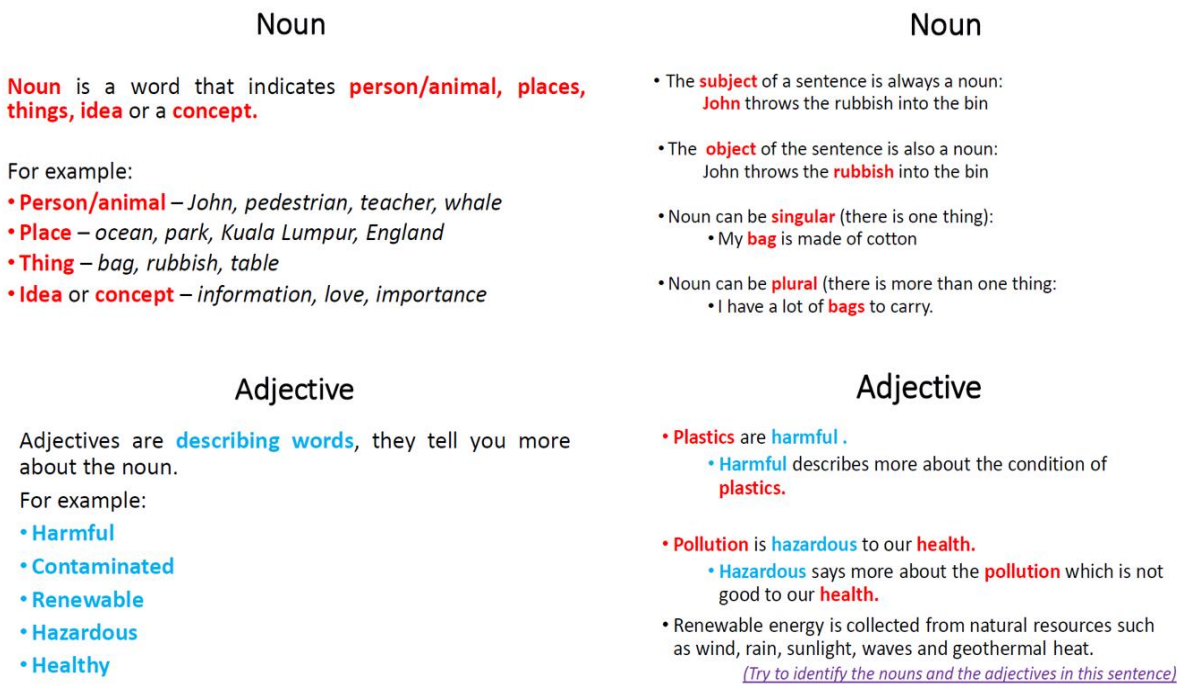

Figure 3. 4: Examples of grammar/vocabulary taught in Lesson 1

In the example of explicit instruction shown in Figure 3.4, the definition of noun and adjective were explained to the students. The grammar and vocabulary items which were taken from Lesson 1 (e.g. environment: pollution), were colour-coded for students to identify them easily. This process in explaining the grammar items and structure was also done in all eight lessons. In addition, vocabulary items that were taught together with the grammar structure in each lesson were related to the respective themes. Examples of how the words function in sentences were also provided, so students could understand and are able to construct on their own. Moreover, the researcher also asked students to identify and make their own sentences using the grammar structure and lexical items. Besides that, the researcher also provided grammar/vocabulary exercises as a follow-up activity after the forms and structures had been explicitly taught. These short grammar/vocabulary exercises were a combination of one to three activities which included either classifying, circle/underlying, fill-in-the-blanks/cloze passage, error identification or sentence completion. The combination of exercises was employed in all eight lessons and varied in each lesson.

3.5.2.2 Instruction for the ST Group

According to Ellis (2009b), the pre-task stage concerns various activities that the teacher and students can undertake to prepare students to perform the actual task that will promote acquisition. As the ST group received TBLT, explicit instruction was not given prior to the speaking tasks (i.e. in the pre-task stage). Instead, students in the ST group were shown a short video related to the theme/topic (e.g. Lesson 1 – environment: pollution). Follow-up questions and a brainstorming activity were conducted to elicit students' interest and thinking towards the theme/topic introduced. These activities were conducted to prepare students before the actual speaking tasks.

In order to distinguish the teaching and learning process for the two groups, Table 3.2 below illustrates the features of the eight-week intervention. According to Ellis (2019), focused tasks (see Section 3.5.3.1) are required in both TSLT and TBLT. However, for TSLT, focused tasks are designed to create contexts for the use of predetermined linguistic features which have been explicitly taught before the actual task while trying to achieve a communicative outcome. Meanwhile, for TBLT, there is no attempt to make students aware of the target features before performing the task and attention to the target forms arises incidentally while performing. Following the definition of Ellis (2019), the differences between the groups resided primarily within the activities conducted in the pre-task stage, making one the TSLT (EIST group) and one TBLT (ST group).

Table 3. 2: *The differences between the EIST group and the ST group*

Features		**EIST Group**	**ST Group**
Syllabus		The same syllabus was employed in both groups	
Approach		Task-supported language teaching	Task-based language teaching
Stages of Teaching	Pre-task	• Introduced the topic/theme by showing video • Explicit instruction • Grammar/vocabulary exercises	• Introduced the topic/theme by showing a video • Brainstorming, question and answer relating to the theme/topic
	During-task	• Pair and/or group discussion which involved decision-making activities • Group presentation	Similar to the EIST group
	Post-task	Students' presentations were discussed in the classroom as a whole and feedbacks were given related to presentation skills	Similar to the EIST group
Teacher's role		Assisted students when needed throughout the speaking tasks	Similar to the EIST group
Students' role		• Active engagement in communication • Opportunities were given to negotiate meaning during speaking tasks	Similar to the EIST group

In addition to the instruction, feedback was also given to both groups. Feedback was given as it is considered as one of the FoEL course requirements, in which all teachers or lecturers are required to provide feedback on students' language performance (e.g. written assignments, oral presentation etc.). In the during-task stage, feedback in both groups was the same where the teacher (i.e. researcher) provided confirmation checks (e.g. *Are you sure? What? You need to double-check?)* while observing students' group discussions and explicit correction was not given. However, following the course requirement, explicit correction was given in the post-task stage after all group presentations. For the purpose of this study, the explicit correction on students' presentation was not given rigorously to avoid overshadowing the effects of the

instruction employed. Errors were not directly pointed out, but instead were randomly selected from several group presentation.

3.5.3 Speaking Tasks in the Intervention

The speaking tasks were designed based on the reviewed literature of task characteristics selected in TBLT research. The nature of the speaking tasks and how they were designed may facilitate different types of interactions and produce different linguistic outcomes. These task characteristics were identified and considered for the purpose of the study.

3.5.3.1 Unfocused versus Focused Tasks

Tasks can be employed in the form of two types, unfocused and focused tasks. The difference between the two types is based on how they are designed. Ellis (2003) explained that unfocused tasks aim to induce general usage of the language without introducing specific linguistic features (e.g. talk about a significant other which involves negotiations between students to complete the task), whereas focused tasks are communicative tasks that may predispose students to process a specific predetermined linguistic feature, receptively or productively. Ellis and Shintani (2014) described a focused task as a "situational grammar exercise" (p. 138) in which the linguistic features are not explicitly taught but the task is designed in such a way that it can only be achieved if learners use specific linguistic features. For example, a narrative task on a past event which requires students to describe and retell their experience in the past. In order to achieve the task, students may have to use past tenses. Although focused and unfocused tasks are different, Ellis (2003) further explained that they must meet the main principle for a task, in which tasks must have some kind of gap (e.g. information or opinion) and meaning-negotiation activities to allow students to draw upon their own linguistic and non-linguistic resources to complete the communicative task (Ellis & Shintani, 2014). Previous studies have shown that focused tasks elicit students to use the target structure more effectively

rather than unfocused tasks which resulted in positive impact on students' retention of new L2 vocabulary (de la Fuente, 2006) and grammar achievement (Ahour & Shemshadsara, 2015).

3.5.3.2 One-way versus Two-way Tasks

Information flows among students could either be one-way or two-way interaction. Ellis and Shintani (2014) described that one-way tasks occur when one student has the information that needs to be communicated to another student. For example, making a hotel reservation which requires one student to act as the speaker (i.e. customer) and the other as the listener (i.e. hotel receptionist) (Nunan, 2004). Meanwhile, two-way tasks involve exchanging information between two interlocutors in order to achieve task completion (Ellis & Shintani, 2014). According to Doughty and Pica (1986), information tasks promote more meaning negotiation among students and facilitates second language acquisition as they allow comprehensive input during interaction. For example, a jigsaw task where students are provided with different information (e.g. two different locations) and are asked to interact with each other in order to find the exact location of a building (Mackey, 2012).

3.5.3.3 Convergent versus Divergent Tasks

Every task incorporates different goals which may be viewed as convergent or divergent (Nunan, 2004). Convergent tasks require students to agree on a single solution and come to a certain consensus before task completion, whereas divergent tasks encourage a variety of responses and students are allowed to disagree and defend opposing views (Ellis, 2003; Nunan, 2004). Mackey (2012) suggested an example of a convergent task such as an expedition to the moon. In this task, students are given a list of items and they are required to select only five items within a 20-minute time before the spaceship leaves. The task encourages students to discuss among themselves to reach a final agreement. However, for divergent tasks, tasks are open-ended and any outcome is acceptable with justification. For example, conducting a debate about whether social media among children brings more harm than good. In this task, students

may provide different responses whether to agree or disagree. The task promotes intellectual discussion where students defend their standpoint or refute their opponents' position.

3.5.3.4 Complex versus Simple Tasks

The complexity of the task may influence students' linguistic outcomes. This is because linguistically complex and demanding tasks produce more uptake of information and more attention to form compared to less complex tasks (Robinson, 2001). The complexity of the task differs according to three main aspects namely: code complexity (e.g. lexical diversity and linguistic complexity); cognitive complexity (e.g. topic familiarity and cognitive processes such as sufficiency of information and information organisation); and communicative stress (e.g. time pressure, number of participants (Skehan, 1998; Skehan & Foster, 2007). Therefore, the nature of the complex tasks can increase students' linguistic production as it requires them to be involved in turn taking and clarification requests, hence more interaction (Robinson & Gilabert, 2007).

According to the reviewed literature on task characteristics, the eight speaking lessons were designed as the present study employed the following characteristics:

- Focused tasks – speaking tasks in the present study were designed to elicit students' use of linguistic features.
- Two-way tasks – reciprocal two-way tasks were used to promote information exchange among students.
- Both convergent and divergent – both tasks required students to reason and make decisions. Convergent tasks encourage teamwork and group interaction, whereas divergent tasks encourage intellectual discussion among students.
- Complex tasks – as the speaking tasks employed decision-making tasks it was important that tasks were cognitively demanding.

3.5.4 Speaking Lessons

After careful consideration of the issues which were discussed in the previous sections, details of the eight-week speaking lessons are explained. All lessons throughout the intervention followed the same stages and procedures; although, the activities completed within a given session (in particular the grammar/vocabulary exercises, see Section 3.5.2.1) varied from lesson to lesson. Therefore, to illustrate the procedures followed in each lesson, Lesson 1 is explained in detail, and Lessons 2 to 8 are summarised as below. A detailed example of the lesson and activities are presented in Appendix A (Lesson 1).

Lesson 1: Environment – *Save our planet!*

In the first lesson, the aim was to explain nouns and adjectives, and introduce vocabulary related to pollution, as shown in Figure 3.4 (highlighted in blue and red). In both groups, the lesson started off with a short video taken from YouTube on the overuse of plastic bags and the negative effects it can bring for marine animals. In the pre-task stage, the video was shown and students in the EIST group were required to focus on the language used in the video (e.g. nouns and adjectives). The teacher (i.e. the researcher) then asked students to list the words verbally mentioned in the video, before providing an explicit definition of the word classes-noun and adjective. A grammar/vocabulary exercise (i.e. identifying nouns and adjectives from an audio and filling the blank with the missing vocabulary) was given once the teacher had finished explaining the target structure.

As for the ST group, the same material was used (i.e. the YouTube video) but the students were instructed to focus on the content of the video (i.e. observe the storyline or the plot of the video). Next, teacher asked questions related to the video (e.g. *what was it about? why did it happen? who were affected?*) and students were asked to brainstorm. In the during-task stage, both groups were asked to form small groups of 3 to 4 students. They were required to discuss

how to avoid or stop pollution and make the world a greener place. Each group was told to come to a consensus on what is the best solution and present their ideas in front of the class. In the EIST group, the teacher reminded the students to use the grammar and vocabulary items they had identified in the pre-task stage during the group activities. Finally, in the post-task stage, the teacher recapped the topic and provided feedback on the students' presentations.

Lesson 2: Sports – *I like to play that too!*

In the second lesson, the aim was to explain verbs and adverbs, and introduce vocabulary items related to sports.

1. Pre-task stage: A short video from YouTube on famous sports in the Olympics was shown to both groups. Grammar/vocabulary exercises were given to the EIST group (cloze passage and error identification), whereas the ST group was asked to brainstorm about the video.
2. During-task stage: Both groups were given several pictures which showed different types of sports in the Olympics (e.g. tennis, football, gymnastics, hockey etc.). Students were required to become a sport analyst in which they had to describe one of the sport events, followed by group presentations. The EIST group was reminded to use verbs and adverbs when describing the sport event.
3. Post-task stage: The teacher concluded the topic and provided feedback on the students' presentations.

Lesson 3: Education – *Studying abroad or just local?*

In the third lesson, the aim was to explain pronouns (personal, reflexive, possessive) and introduce vocabulary items related to education.

1. Pre-task stage: A short video from YouTube on experience of studying abroad was shown to both groups. Grammar/vocabulary exercises were given to the EIST group

(e.g. classifying pronouns and cloze passage). The ST group was asked to brainstorm about the video.

2. During-task stage: Both groups were required to discuss whether studying abroad is better than studying locally, followed by group presentations. The EIST group was reminded to use pronouns in their group activities.
3. Post-task stage: The teacher summarised the topic and gave feedback on the students' presentations.

Lesson 4: People and Lifestyles – *Crazy things people do!*

In the fourth lesson, the aim was to explain articles and determiners and introduce vocabulary items related to people and lifestyles.

1. Pre-task stage: A short video from YouTube on crazy habits that people do to make themselves beautiful was shown to both groups. A grammar/vocabulary exercises were given to the EIST group (e.g. classifying, fill-in-the-blanks/cloze passage and error identification), whereas the ST group was asked to brainstorm about the video.
2. During-task stage: Both groups were required to conduct a debate on the advantages and disadvantages of plastic surgery and how it affects our lives, followed by group presentations. The EIST group was told to use article and determiners throughout the class.
3. Post-task stage: The teacher concluded the topic and provided feedback on students' presentations.

Lesson 5: Leisure and Pastimes – *Oops what did I do?*

In the fifth lesson, the aim was to explain the simple present and past tense, and introduce vocabulary items on leisure and pastimes.

1. Pre-task stage: A short video from YouTube on a journey of a successful crippled man was shown to both groups. Grammar/vocabulary exercise was given to the EIST group

(e.g. circle the answer, sentence completion and error identification). The ST group was asked to brainstorm about the video.

2. During-task stage: Both groups were required to discuss a life experience that they had in the past and whether to regret it or embrace it, followed by group presentations. The EIST group was encouraged to include simple present and past tense in their discussion.
3. Post-task stage: The teacher concluded the topic and provided feedback on students' presentations.

Lesson 6: Tourism and Travel – *I wish to go there!*

The sixth lesson was a continuation of the previous lesson. The aim was to explain the present and past continuous tense, and the present and past perfect tense, as well as to introduce vocabulary items on tourism and pastimes.

1. Pre-task stage: A short video from YouTube on traveling to amazing places was shown to both groups. Grammar/vocabulary exercises were given to the EIST group (e.g. fill-in-the-blanks/cloze passage and sentence completion). The ST group was asked to brainstorm about the video.
2. During-task stage: Both groups were required to plan a destination to go to, i.e. one of their favourite places in the world, and what will they bring with them, followed by group presentations. The EIST group was advised to use the grammar items discussed in the pre-task stage.
3. Post-task stage: The teacher summarised the topic and commented on students' presentations.

Lesson 7: Health – *I should have…*

In the seventh lesson, the aim was to explain modal verbs, and introduce vocabulary items on health.

1. Pre-task stage: A short video from YouTube on healthy lifestyle and the best way to achieve it was shown to both groups. Grammar/vocabulary exercises were given to the EIST group (e.g. sentence completion and error identification). The ST group was asked to brainstorm about the video.
2. During-task stage: Both groups were required to discuss the do's and don'ts in achieving a healthy life and which was the best way, followed by group presentations. The EIST group was asked to use modal verbs in their activities.
3. Post-task stage: The teacher summarised the topic and gave feedback on students' presentations.

Lesson 8: Science and Technology – *See you in the future!*

In the eighth lesson, the aim was to explain the future tense, and introduce vocabulary items on science and technology.

1. Pre-task stage: A short video from YouTube on the evolution of gadgets was shown to both groups. Grammar/vocabulary exercises were given to the EIST group (e.g. fill-in-the-blanks/cloze passage, sentence completion and error identification). The ST group was asked to brainstorm about the video.
2. During-task stage: Both groups were required to choose a technological device (e.g. mobile phone, laptop, tablets etc.) and predict the changes that will happen with the chosen device after 50 years. Students needed to discuss the changes and the benefits of the future device and present their ideas in groups. The EIST group was reminded to include future tense in their group activities.

3. Post-task stage: The teacher concluded the topic and provided feedback on students' presentations.

3.6 Instruments

In this section, the instruments used for data collection are described. Four instruments were used in the present study which are vocabulary levels test, speaking test, questionnaire and semi-structured interview. The latter three were conducted in both pre- and post-test stage but the vocabulary levels test was only administered in the pre-test stage.

3.6.1 Vocabulary Levels Test

As mentioned in Section 3.3 and 3.4, students were randomly assigned into two experimental groups. Due to randomisation, a general vocabulary knowledge test was employed in this study in order to provide additional evidence of the students' English language proficiency level before the intervention. Therefore, the present study adopted a vocabulary levels test (VLT) designed by Schmitt, Schmitt, and Clapham (2001) which can be obtained from the website, (printable and online, http://www.lextutor.ca/tests/). The VLT is designed to measure students' vocabulary size of general and academic English because high vocabulary size reflects the learner's ability to use the language in various ways (Kremmel & Schmidt, 2018; Schmitt et al., 2001). Following Schmitt et al. (2001), the VLT in this study had five sections to measure students' levels of English word families which included 2000, 3000, 5000, 10000 and academic. For each section, there were a total of 10 groups of clustered words. Each group contained six words on a left column and combined with three items of the corresponding meaning on the right column in a format of multiple matching, as shown in Figure 3.5 below.

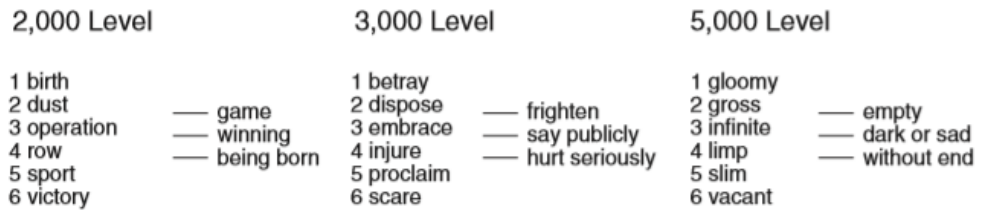

Figure 3. 5: Examples from the vocabulary levels test (Schmitt et al., 2001)

The three items in the right column (30 in total in each section) were used to represent 100 words of the particular frequency level (i.e. 2000, 3000, 5000, 10000 and academic). Within each level, the vocabulary were distributed based on a "fixed ratio of word classes to represent the distribution of English word classes" (Kremmel & Schmidt, 2018, p. 1). This ratio was 3 (noun): 2 (verb): 1 (adjective) which is equivalent to 15 nouns, 10 verbs and 5 adjectives in each section. A full version of the VLT is presented in Appendix B.

3.6.2 Speaking Test

Language assessment is used as a tool to enable teachers or researchers to make inferences about students' language ability and their performance on tasks (Bachman & Palmer, 2010). As for this study, the speaking test was utilised to elicit students' speaking performance in terms of complexity, accuracy and fluency in order to observe the effect of different instruction during the intervention. To capture overall speaking performance, several constraints should be taken into consideration as different tasks elicit different cognitive load and this may influence language outcomes and the overall performance on the task (Foster & Skehan, 1996).

Nakamura (2009) presented a three-dimensional approach to speaking assessment tasks namely monologue, dialogue or multilogue which focus on the number of people involved in the process. Monologue is a one-way speaking activity that can range from a simple phonetic level check (pre-monologue) to advanced oral presentation skills (creative monologue) where no listening skill is required. Dialogue, on the other hand, refers to a two-way speaking activity which involves the ability to exchange information in a situation where the context is well

defined and fixed in speech e.g. classroom (transactional), and the ability to perform with consideration to deeper predictable social relationships (interpersonal). Meanwhile, multilogue is identified as discussion among multiple participants and advance listening skills are required for this activity. In addition, dialogue and multilogue are considered more difficult and complex than monologue as they can be argumentative and involve various communication strategies (Nakamura, 2009).

However, monologue and dialogue are the types of tasks that are frequently used in measuring oral language. Although it is arguable that a dialogue allows for authenticity and is characterised as interactive by nature (Guillot, 1999; van Lier, 2004), monologic performance elicited by tasks has been frequently employed in numerous studies, for example picture description (Albino, 2017; Teng, 2007), oral narrative (Awwad, Tavakoli, & Wright, 2017; Qiu, 2019; Tavakoli, 2011), short talk (de Jong & Perfetti, 2011), oral presentation based on word description (Teng, 2007). The frequent use of monologues is due to a number of factors, as suggested by Tavakoli (2016, p. 3):

- the degree of control associated with a monologic task performance (i.e. simpler pragmatic demands for speech planning),
- predictability of the outcome of the performance, and
- clarity and ease of the procedures for measuring language produced in a monologic task.

Dialogue, in contrast, has been shown to be difficult as it involves speakers taking turns to talk which leads to complex pragmatics (Tavakoli, 2016). Therefore, assessing a dialogue could be a challenge as it involves the interactive aspects of dialogue such as "overlap, unclaimed between-turn pauses, and the interdependence of the interlocutors' performances" (Tavakoli, 2016, p. 3). In other words, dialogue tasks require the interaction of both speakers in conveying meaning and throughout this process, advanced listening and communication strategies are

needed. However, Michel, Kuiken, and Vedder (2012) argued that dialogue is less demanding compared to monologue not only because the interaction between speakers is easier to engage in, but because speakers have the opportunity to use the interlocutor's turn to plan and think about their own subsequent performance. That said, a speaker's utterance could be another speaker's idea and hence speakers might lack of originality and creativity in producing speech.

Therefore, due to the nature of the monologue mentioned by Tavakoli (2016) and Nakamura (2009), the present study decided to employ creative monologue for both pre- and post-test which was also in line with the Task A of the MUET speaking test. By using a monologic task in the speaking test, it allowed comprehensible assessment of each student's English speaking performance without the influence of other variables such as the influence of group interaction.

In an attempt to reduce the potential bias of unfamiliarity with the speaking task design, the study adapted the MUET speaking test as students are familiar with the test format. The MUET speaking test has two different tasks. Students are required to perform two tasks: an individual presentation and group interaction as shown in Table 3.3 below. Students are assigned into a group of four and will be determined as candidate A, B, C and D by the examiner. For the individual presentation, each candidate is required to give an opinion based on the task given in which each individual task has a different situation. The situation given is usually related to contemporary issues such as personal experience, education, health, sports, science and technology, etc. However, in this task, students are not required to make any decision and only asked to present their ideas and describe based on the information given to them. Two minutes will be given to prepare and later two minutes for each individual presentation, from candidate A to candidate D. While one candidate is presenting, other candidates are required to listen and take notes which later leads to decision-making in the group interaction.

Table 3. 3: *Details of the tasks in the MUET speaking test*

Task		Participant	Task Characteristics		
			Dimension	**Condition**	**Type**
A	Individual presentation	One student	Monologue	Preparation – 2 minutes Presentation – 2 minutes	Divergent
B	Group interaction	Four students	Dialogue	Preparation – 2 minutes Discussion – 10 minutes	Convergent

Meanwhile, for the group interaction task, each candidate is given two minutes to prepare points to support or oppose the other candidates' views based on the previous task. Then, 10 minutes will be given for the group interaction in which they have to argue with each other and come to a consensus before the time ends. Due to the nature of the monologue task (i.e. describing) and the dialogue task (decision-making) in MUET, the present study decided to adapt both characteristics by combining the monologue task and decision-making task together. In addition, several keywords or prompts were provided to the students in both speaking tests (i.e. planning time prior to the speaking tests) to elicit response and direct students to towards the topic of discussion (Cox, 2013) to prevent students plenty of leeway on what to talk about during the speaking tests and to ensure they remain within the topic of discussion. The pre- and post-speaking tests are presented in Appendix C.

3.6.3 Questionnaire

An important aim of the study was to investigate how students' motivation towards learning and speaking English specifically evolves over time. The reviewed literature on motivation theory in the previous chapter shows that L2 motivation can be complex in nature and is not static. In order to measure students' motivation, a mixed methods approach was employed in this study in which the methodology combined both quantitative and qualitative data. By using

a mixed methods approach, it allows a more comprehensive assessment of students' motivation. In this study, quantitative data were collected first through a questionnaire and then followed by qualitative data through an interview. The combination of both questionnaire and interview allowed the study to "measure the changes in motivational intensity (scales mean scores) and uncover the motivational thinking behind the attitudes thus enhances the inherently superficial questionnaire data" (Courtney, 2014, p. 134). The following sections describes the instruments used in the study.

Questionnaires are widely used for collecting survey information and for providing structured and mostly numerical data. The main purpose of questionnaires is to seek factual information that involves background and biographical information, knowledge and behavioural information as well to measure attitudes, values, opinions and beliefs (Punch, 2009). For the present study, a questionnaire was designed and developed to evaluate students' L2 motivation towards the English language and speaking English. The questionnaire at the pre- and post-test contained similar items but met different purposes. The questionnaire in the pre-test was administered to obtain information about students' past experience before the intervention, whereas the post-test questionnaire was to gain insights into students' motivation and experience during the intervention period. The questionnaire presented in Appendix D contained two parts which are outlined in Table 3.4 below.

Table 3. 4*: Outline of the pre/post-test questionnaire*

Part	**Details**	**No. of Items**
A	Demographic information	7
B	Motivation towards the English language and speaking English	45
	Total	**52**

Part A contained six items about students' demographic background such as age, gender, ethnicity, MUET result, frequency in speaking English, rating of their overall English speaking performance. These demographic data were considered important for the study as they

provided a better understanding of the population size, distribution and composition (Murdock & Ellis, 1991). In other words, the demographic variables categorise overall survey response into meaningful group of respondents and give a clear perspective on the respondents' background.

As for part B, questionnaire items were based on an instrument used by Zainol Abidin, Pour-Mohammadi, and Alzwari (2012) which was adapted for an ESL context from the Attitude/Motivation Test Battery (AMTB) (Gardner, 1985). The reason for using the AMTB is that it has been widely-used by various studies of L2 motivation and empirically well tested. Originally, questionnaire items in Zainol Abidin et al. (2012) comprised a total of 45 items. Of these,15 items were adopted for the present study in an unchanged format. However, some other items had to be revised for clarity reasons and to meet the purpose of the study which specifically focuses on L2 motivation towards English language and speaking English. These modified items, which were 22 in total, were either added to with a new word, reworded and/or modified by simplifying the sentence structure. A further eight items from the previous study were disregarded as they were considered not related to the present study. To replace these items, eight new ones were self-developed by the researcher based on the reviewed literature of L2 motivation (Dörnyei, 1994) and the researcher's experience in teaching English. Moreover, the items were selected in accordance with Dörnyei's (1994) L2 motivation framework which enabled a multi-faceted view of the development of students' motivation over time rather than measuring motivation as a whole or as a one concept only. A few examples of the items that were adopted and modified from Zainol Abidin et al. (2012) and self-developed by the researcher based on the reviwed literature are shown in Table 3.5 below.

Table 3. 5*: Examples of the adopted, modified and self-developed items*

Items in the Pre/Post-Questionnaire		Original items from Zainol Abidin et al. (2012)		Changes
No.	Items	No.	Items	
1	I look forward to the time I spend in English class.	44	I look forward to the time I spend in English class.	U
2	Speaking English is important because it will make me more educated.	1	Studying English is important because it will make me more educated.	M
3	I do not feel enthusiastic to come to class when English is being taught.	42	I do not feel enthusiastic to come to class when the English is being taught.	U
18	I am interested in doing speaking activities in English.	28	I am interested in studying English.	M
33	Speaking English gives me opportunities to meet new people.		–	SD
45	I can be successful in work and life if I speak good English.		–	SD

Notes: U=Unmodified; M=Modified; SD=self-developed

These items were categorised into five main motivation scales on three level components in line with the proposed motivation framework (Dörnyei, 1994) presented in Table 3.6 below.

Table 3. 6*: Number of items according to the motivation sub-scales*

Components	Scales	Items	No. of Items
Language Level	Attitude	22, 24, 25, 34, 37, 40, 43	7
	Integrativeness	9, 26, 30, 33	4
	Instrumentality	2, 5, 12, 13, 17, 19, 41	7
Language Level	Linguistic Self-confidence	8, 10, 14, 15, 16, 20, 21, 23, 27, 28, 29, 31, 32, 35, 36, 38, 42, 44, 45	19
Learning Situation		1, 3, 4, 6, 7, 11, 18, 39	8
		Total items	**45**

The attitude scale included items related to students' overall attitudes towards the English language and willingness to communicate in English. For the integrativeness scale, the items

focused on students' attitudes to the target language community and communicating with target language speakers. The instrumentality scales incorporated items related to the purpose and usefulness of learning English language. The items for linguistic self-confidence concentrated on speaking anxiety, self-confidence and self-efficacy associated with the students' perception of their English speaking performance. Finally, the learning situation scales concentrated on extrinsic factors such as the role of the teacher and teaching methods used in the classroom and speaking task motivation.

In the actual questionnaire, these items were mixed up together in order "to create a sense of variety and avoid respondents from simply repeating previous answers" (Dornyei & Taguchi, 2010, p. 47). A structured questionnaire was designed to enable patterns to be observed and comparisons to be made (Cohen et al., 2018). Therefore, respondents have no leeway in term of a personalised input and all responses fall into certain categories (Newby, 2014). In addition, a four-point Likert scale ranging from *strongly disagree (1), disagree (2), agree (3)* and *strongly agree (4)* was employed in the questionnaire. The mid-point or the *undecided* option was omitted because "when a mid-point is inserted, there is evidence that responses can gravitate towards it and there is comfort in the average and it is easier to check it rather than think deeply about the issue and decide on which side to sit" (Newby, 2014, p. 308). In other words, students had to decide which side reflect their attitude and motivation the most, whether it was positive or negative towards the English language and speaking English.

3.6.4 Semi-structured Interview

Interview is a method most often used as a mean of obtaining in-depth information on individuals' experiences, attitudes, perceptions, beliefs, thoughts, knowledge and feelings of a problem being researched (Dornyei, 2007; Talmy, 2010). Therefore, through interview, it allows the researcher to elicit a greater understanding of the key concepts of students' L2

motivation towards the English language and speaking English. In this study, a relatively open interview format involving a semi-structured interview was employed as it is open-ended which allows for probing, follow-up and clarification (Dornyei, 2007). The items of the interview were developed according to the same theoretical framework as the questionnaire. This was because the interview was to supplement the existing questionnaire and thus to gain a deeper insight and allow detailed discussion of the issues pertaining to students' L2 motivation and language learning experience. Similar to the questionnaire, the same semi-structured interview was used in different times (pre and post-test) to meet different purposes. The interview presented in Appendix E contained 15 main questions in relation to the five scales of motivation that aimed to elicit students' L2 motivation in depth. Follow-up questions were also used to obtain additional information if the students provided vague and obscure responses. Moreover, the questions were kept short, direct and jargon free in order to obtain authentic responses dues to the time constraint as students were from different majors and had different schedules.

3.7 Pilot Study

Pilot study refers to a small-scale trial of a study conducted before the full-scale trial (Gay et al., 2011). The primary purpose of piloting is to ensure that the research instruments could be used effectively in the actual study (Bryman, 2012). In addition, piloting helps to easily identify items that contain ambiguity or are unlikely to be of interest within the study as they might not form a variable (Bryman, 2012). Therefore, it is essential that the instruments are piloted in order to establish content validity (Creswell, 2014). For the present study, the pilot study involved administering the questionnaire, the speaking test and the interview, after Ethical Approval was obtained (see Section 3.9) and informed consents were gained from the participants.

3.7.1 Questionnaire

The questionnaire was piloted several months before the actual study. The participants were selected through convenience sampling to enable a higher response rate (Bryman, 2012). Cohen et al. (2018) suggested that gathering sufficiently large-scale data allows the researcher to calculate "reliability levels (alphas), item difficulty and item discriminability, and identify commonly misunderstood or non-completed items" (p. 583) more easily. Therefore, 115 students were recruited for piloting the questionnaire. These students were asked to answer the questionnaire via the Google Forms application; the link to the questionnaire was sent through their social media such as Facebook and Twitter. To maximise the acceptability of the questionnaire, reliability testing was used. Reliability refers to the degree of consistency of what is being measured (Gay et al., 2011). Reliability was assessed by measuring internal consistency using SPSS version 24 to calculate Cronbach' Alpha Coefficient. Cohen et al. (2018) suggested that a scale is considered reliable and acceptable when $\alpha > .70$. The reliability test result showed that the questionnaire was reliable and appeared to have good internal consistency, $\alpha = .905$. All the items were worthy of retention and no changes were made after piloting. Although the greatest increase in α would come from deleting item 3 ($\alpha = .907$), and item 14, 16 and 17 ($\alpha = .906$), removal of these items would increase the alpha only by .002 or .001. This increase is not dramatic and the items were considered crucial to be asked in the questionnaire. Therefore, the alpha value reflected a reasonable degree of reliability.

3.7.2 Speaking Test

Two sets of the speaking test which consisted of two questions were adapted from the previous MUET speaking test. The two sets were obtained directly from an English teacher who had experience as both MUET teacher and examiner for more than 10 years. Although the previous MUET speaking test had been vetted by experts, piloting was considered necessary for this

study as some of the items were altered to suit it. The two sets were Set A, self-reported or personally related to students' experience (e.g. how to overcome tiredness and which is the best way overcome tiredness) and Set B, where questions were more factual and cognitive demanding (e.g. major causes of climate change and how to overcome it, and reasons children should be banned from gadgets). However, both sets were allocated the same time length and were the same type of decision-making tasks in which students had to describe and justify their decision. Both sets were piloted by the researcher's colleague in Malaysia on four ESL undergraduate students. Their speaking tests were audio-recorded and sent to the researcher to be analysed. Students were also asked to give feedback on the two sets of speaking test.

Students were able to comprehend and answer both items. However, the major concerns were the length of the test and the language produced. In set A, the results showed that students were able to produce approximately 350-450 tokens. For Set B, however, the results showed that students produced approximately 200-300 tokens with a high production of disfluency. As for the students' feedback, the results showed that both sets were comprehensible and they understood each question clearly. However, students mentioned that the time length given for each test was insufficient especially in Set B where the questions were more factual, critical, demanding and required a high level of general knowledge. The present study investigates the effects of two different forms of language teaching (TBLT and TSLT) on students' speaking performance rather than the effects of task complexity, hence Set A was selected for the pre- and post-speaking test rather than Set B. Moreover, Set A also provided richer data of a sufficient amount to be analysed.

3.7.3 Semi-structured Interview

The semi-structured interview was piloted with three undergraduate ESL students by the researcher himself a month prior to the actual study. During the interview, students were informed about the

interview and that it would be audio-recorded. The interview lasted approximately 40-45 minutes. The researcher then listened to the audio and transcribed the interview. A few questions were found to be ambiguous which led students to hesitate and to answer repetitively which resulted the interview to be lengthy. Therefore, the researcher decided to disregard these questions. Questions that prompted students to elaborate their answers were kept. Only a few questions needed to be reworded as they were found to be unclear to the students.

3.7.4 Speaking Intervention

The intervention was piloted to evaluate its suitability and practicality before the actual study. It worth noting that only one lesson was piloted due to limited time available and also the researcher's availability. One teacher, who is the researcher's colleague, agreed to conduct the pilot study. A copy of one of the lesson plans and also feedback forms for students and the teacher were given via email to the teacher to be employed in a Malaysian university setting. Before the pilot study, the researcher explained the procedure to the teacher and requested that the classroom be video recorded. Through convenience sampling, 24 ESL undergraduate engineering/science students who were enrolled in the FoEL course were selected to participate in the pilot study. These students were current students in one of the classes taught by the teacher herself. The researcher received the video-recoding and feedbacks from both students and teacher two weeks later. Based on the observation from the video, students reacted positively towards doing the speaking tasks. The speaking tasks were conducted in groups of four. Students were able to participate and interact according to the demands of the task. Following the observation, the majority of the students gave positive feedback towards the speaking tasks. However, the teacher commented that pair work was not conducted frequently due to limited time and group work was considered sufficient. In addition, the teacher commented that due to time constraints, it was impossible to do individual presentations.

Accordingly, the results of the pilot study indicated that the lesson was feasible and workable with pair/group work discussion and group presentation only.

3.8 Data Collection Procedures of the Main Study

The process of data collection commenced by distributing the questionnaire and was immediately followed by the individual speaking test. The interview was conducted last, once after all quantitative data had been collected. The pre-test data collection was conducted during the first two weeks before the intervention. Meanwhile, the post-test data collection was conducted immediately once the intervention ended and before the final examination week.

3.8.1 Vocabulary Levels Test

The vocabulary levels test was administered on the first day in the first pre-test week after the researcher had explained the purpose of the study and gone through the information sheet and the consent form with the students (see Section 3.9 for ethical procedures). Once the students understood the purpose of the study, the researcher collected their consent forms and immediately students were given the vocabulary levels test. The researcher explained how to answer the vocabulary test by showing one example. The test took approximately 15 to 20 minutes to complete.

3.8.2 Questionnaire

Similarly to the vocabulary levels test, the pre-test was administered after students have answered the vocabulary test. This was done on the same day for students to reflect on their motivation and past experiences in learning and speaking English, and also to avoid the influence of variables such as the teacher (i.e. the researcher) and the intervention. Before administering the questionnaire, the researcher explained that students' anonymity would be kept safe and confidential. Then the researcher explained all items thoroughly to ensure that all

students understood before making a choice and also to ensure a 100% response rate. Instructions were also provided in the questionnaire at the beginning of each section. The questionnaire took approximately five to 10 minutes and was collected immediately by the researcher. A similar process was undertaken for the post-test questionnaire which was conducted immediately after the intervention to minimise the gap between the intervention and the post-test questionnaire. Students were asked to reflect on their motivation towards the learning the English language and speaking English during the intervention week.

3.8.3 Speaking Test

The speaking pre-and post-tests were conducted immediately after the pre-test questionnaire and intervention, respectively. Unlike reading and writing tests which can be administered in groups, a speaking test is usually conducted individually. The speaking pre- and post-test were conducted over nearly two weeks because students were in different courses and they had different time schedules, thus it was impossible to conduct the speaking tests in one session.

It is worth noting that planning time can influence students' language outcomes. Generally, planning time should offer students ample time for task performance (Bui, Skehan, & Wang, 2018). Planning as described by R. Ellis (2005b) is "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). There are two types of planning: *pre-task planning* and *within-task planning*. As for the former, R. Ellis (2005b) further explained that it provides students with an opportunity to perform the task before the actual performance. Pre-task planning can be differentiated into *rehearsal*, which entails task repetition before the main performance; and *strategic planning*, which involves preparation before the main performance by considering the content that needs to be encoded and expressed (R. Ellis, 2005b). Numerous studies have shown that strategic planning was effective to elicit students' language outcomes. Most studies employed 10

minutes for planning time (Mochizuki & Ortega, 2008; Pang & Skehan, 2014; Van de Guchte et al., 2017), while others adopted less planning time such as five minutes (Tavakoli & Skehan, 2005), three minutes (Elder & Iwashita, 2005), one minute (Wigglesworth, 1997) or a combination of times (e.g. one, two, three and five minutes) (L. Li, Chen, & Sun, 2015). This indicates that planning time varies across research and is not limited to a specific time frame.

As for the latter, within-task planning, R. Ellis (2005b) distinguished between *unpressured* which involves students engaging in careful on-line planning; and *pressured*, students engage in rapid planning. Both pre-task planning and within-task planning demonstrate distinctive types of planning as shown in Table 3.7 below. R. Ellis (2005b) argued that these should not be identified as "mutually exclusive" (p. 4) and there is a possibility to combine the four types of the two planning conditions together.

Table 3. 7*: Planning conditions (R. Ellis, 2005b, p. 5)*

Planning conditions	Pre-task planning	Unpressured within-task planning
1	No	No
2	Yes	No
3	No	Yes
4	Yes	Yes

The present study was conducted within the FoEL course which was structured based on the MUET specification and format that covered the four skills (e.g. listening, speaking, reading and writing). Similar to condition 2 in Table 3.7 above, the present study decided to adopt the original MUET speaking test format. Students were given strategic planning time prior to the actual speaking test but they were not allowed to plan their speech carefully on-line (i.e. pressured planning). However, an additional 30 seconds were added to the actual speaking tests as shown in Table 3.8 below. This was considered necessary because the questions used in both speaking tests were slightly longer and had some alterations from the original MUET speaking

test, and the results from the pilot study showed that students needed an extra few seconds to complete their final sentences.

Table 3. 8*: Comparison of the planning conditions in both speaking tests*

Planning Conditions	Speaking Tests	
	MUET	**Main Study**
Strategic planning	2 minutes	2 minutes
Pressured within-task planning	2 minutes	2 minutes + 30 seconds

As for the speaking tests procedure, both tests were conducted individually and not during the class in order to avoid interfering with the lesson. The students were notified that their speech would be audio-recorded and used for research purposes only. First, students were given a piece of paper to write down any information needed. Students were reminded not to discuss the question with their classmates once the speaking test had ended and were informed that the paper that was used during the speaking test would be collected. The researcher explained the procedure of the speaking test which also included the format of the test. Students were only allowed to write down keywords instead of writing in sentences to avoid verbatim reciting which would hinder natural speech.

3.8.4 Semi-structured Interview

The researcher selected three students from each group based on their previous MUET scores (e.g. Band 2 and 3). The pre-test interviews were conducted during the second week of the pre-test week once all quantitative data had been collected and before the intervention to avoid influence from the effects of the intervention. Meanwhile, the post-test interviews were conducted within one or two weeks after the intervention in order to minimise the effect of the time gap.

Before students were interviewed, the researcher explained the procedures of the interview. To lessen the feeling of being apprehensive on being audio-recorded, students were notified in advance that this would happen, with their consent, and interview questions were shown to them, so they were well informed of the content before the interview started. This is important to assure that students were not being asked to do something with which they felt uncomfortable. The interview was conducted in both Malay and English so that students felt comfortable when answering and also to encourage authenticity and fluent discourse. As the researcher was also the class teacher, it was important to explain the different roles of the researcher during the interview session and in the classroom to reduce the researcher's sense of authority, thus the interview took place outside the classroom. Each student took approximately 20 to 30 minutes to interview.

Once the interview ended, the researcher played the recorded audio on the computer so that students had the opportunity to listen to what was being recorded in case they wanted to add more information or omit any responses that they felt uncomfortable sharing. All students agreed to the information that they provided during the interview.

3.9 Ethical Considerations

Research in education involves working with human beings as participants and their learning organisations as the location where the collection of data is taken and carried out (Wellington, 2015). In conducting research, the researcher needs to respect participants and the sites involved in the study and also acknowledge that their rights will be protected during the process (Creswell, 2013b). Therefore, official permission from the institution where the research takes place and approval by those who are involved before conducting the research (Cohen et al., 2018) should be taken seriously in any educational research.

In this present study, formal procedures were taken into consideration and were followed strictly. Consent for ethical approval was obtained from the University of Reading, Institute of Education Ethics Committee before any data collection was undertaken. A copy of the signed ethical approval from the committee is presented in Appendix F. Meanwhile, for the Malaysian counterpart, permission to conduct any study in schools, institutions or any other organisations must be obtained from the Economic Planning Unit (EPU), under the Prime Minister's Department of Malaysia which oversees any approvals for any research studies from a foreign university. Permission to carry out research was sought early and the preliminary application was made online through the EPU website (http://www.epu.gov.my). Follow-up by EPU had to be done by furnishing ample information on the research study (see Appendix G for the written consent attained from the EPU).

As this study focuses on the language centre of a technical/engineering university, the institution is under the jurisdiction of the MoE. However, the university is considered as a statutory body and it is independently managed by the university itself. Therefore, approvals between two parties: EPU and the language centre were given and liaised over with the assistance of EPU before any permission was granted for the researcher to conduct the study. An information sheet and consent form were sent to the dean of the language centre stating the period of the study, participants involved in the study and purposes of the study being conducted (see Appendix H for the Information Sheet and consent form for dean of the language centre). Meanwhile, as for the participants of this study, an Information Sheet and consent form were handed out to the students involved before the start of the study. The researcher explained in detail about the study to the students outlining that they would be asked to complete the questionnaires, participate in an eight-week intervention, undertake two speaking tests (pre- and post-test) and would be selected to be interviewed. The Information Sheet listed the name of the researcher, email address and assurance of confidentiality in case

students had any inquires related to the study (see Appendix I for Information Sheet and consent form for students).

3.10 Summary

This chapter discussed the methodology and methods of the present study. The research paradigm and design of the study were outlined followed by the participants involved in this study. The instruments and the procedures were also presented in detail. The pilot study and changes to the main study were also reported and the data collection procedures for the quantitative and qualitative data were also explained. Finally, the procedures followed to ensure ethical conduct in the study were outlined. The following two chapters, Chapter Four and Five presents the data analyses procedures and the findings of the study.

CHAPTER FOUR: ANALYSES & FINDINGS OF L2 SPEAKING TEST

4.1 Introduction

This chapter reports the analyses and results of data that were collected through quantitative methods, namely the pre and post-test which were utilised to answer the main research question one. To reiterate, the first principal research question of the current study is:

1(a). What are the effects of two different forms of language teaching (TBLT and TSLT) on the development of students' L2 speaking performance over time.

In addition, three sub-research questions were developed to investigate students' speaking performance in terms of complexity, accuracy and fluency.

i. How do students' speaking complexity and types of language produced develop over time?
ii. How do students' speaking accuracy and the forms of errors made develop over time?
iii. How does students' speaking fluency change over time?

The objective is to describe the nature of the ESL students' overall speaking performance and assess whether their spoken complexity, accuracy and fluency increased over the semester following either instructional treatment. Speaking data were analysed using two methods:

1. Quantitative, statistical analysis via SPSS and presented first in this chapter; and
2. Qualitative, whereby the types of vocabulary, errors and disfluency produced from both groups over time were ascertained.

4.2 Speaking Data Software Analysis

The speaking data collected from the pre- and post-test were transcribed using Express Scribe Transcription Software. Later, the data were then transferred to another software called Computerized Language Analysis (CLAN) and coded. CLAN is a language analysis software that was developed as part of the Child Language Data Exchange System or CHILDES project (MacWhinney, 2000) and has been widely used in research related to the field of first language acquisition and conversation analysis (Meakins, 2007). Primarily, the software has two components. The first part is the CLAN editor which is used to transcribe and code files in CHAT (Codes for the Human Analysis of Transcripts) and conversation analysis format. The second part of CLAN functions as part of a statistical analysis tool and produces statistical outcomes of the file being transcribed. CHAT can be used distinctly without CLAN but CLAN is reliant on well-coded and transcribed CHAT transcripts. For this study, a total of 118 transcriptions were coded and analysed.

4.3 Lemmatisation of the Speaking Transcriptions and Treatment

Before analysing the speaking data, it is worth noting that the speaking data were lemmatised. As the English language is a highly inflected language, lemmatisation was considered. Lemmatisation involves "replacing all inflected forms of nouns, verbs and adjectives with the corresponding lemma" (Treffers-Daller, 2013, p. 85). One lexeme can form more than one inflected variations of the word. For example, the lexeme, *write* can take many forms such as *write*, *writes*, *writing*, *wrote* and *written*. All of these inflections have the basic meaning denoted by an action that is thus represented by the lemma, *write*. All of the inflections come from the same morphological category, which in this case refers to a verb and there is no difference in the intended meaning. In contrast, the word *writer* does not belong to the same lexeme as the word is not inflected and is derived from a different morphological category,

which is a noun. In some cases, one lexeme can be formed of more than one word and morpheme such as compound words (e.g. write-up, co-write). The meaning of the lexeme is determined by the constituents together and not from the single words taken separately.

Therefore, in this study, to assess students' speaking complexity, only lemmas were counted when analysing students' productive vocabulary. According to Jarvis (2002), using lemmas would avoid confusion between the students' vocabulary knowledge and grammatical knowledge. This can be supported by Vermeer (2004) who argued that inflected forms would mask the learners' lexical knowledge and hence the use of lemmas would be the most valid method for counting lexical richness. Therefore, it would be better to treat a lexeme identically as a lemma, which had been done in past studies (Treffers-Daller, 2013; Treffers-Daller et al., 2016; van Hout & Vermeer, 2007; Vermeer, 2000).

As mentioned above, the transcriptions which contained inflected forms were lemmatised and saved separately, as this aids the statistical count of lexical types of lemma (Jarvis, 2002). The lemmatisation process was completed by referring to the online datasets of lemmatization lists by Mechura (2016) which provided a list of lemmas in various languages including English. Table 4.1 below shows a few examples taken from the transcriptions to illustrate how the data were lemmatised and treated.

Table 4. 1*: Examples of lemmatisation and treatment of the transcriptions*

Categories	Inflection / Lexeme	Lemma / Combination
Articles	a / an	a
Auxiliary Verbs	am / is / are / was / were	be
	be / being / been	be
	have /has / having / had	have
	do / does / doing / did / done	do
Comparative Adjectives	happy / happier	happy
Compound Words	junk food	junk+food
	heart attack	heart+attack
	hang out	hang+out
	dress up	dress+up
Contractions	don't / doesn't / didn't	do not
	I'm / they're / we're	I be / they be / we be
Formulaic Phrases	on the other hand	on+the+other+hand
	last but not least	last+but+not+least
Plurality	feelings	feeling
	these / those	this / that
Tenses	contribute / contributed	contribute

However, in some cases, an English word can carry more than one meaning, which the CLAN software cannot automatically distinguish. As an example, in the present study, the words most frequently uttered by the students were *study* and *play*. These words can be either a verb or a noun depending on how they are used in context. To allow CLAN to differentiate between the two, the original form would be considered as a verb, and where it occurred as a noun, *+n* was added at the end of the word (e.g. *study+n* and *play*+n). Therefore, CLAN would recognize both words as separate types (useful for the measure of lexical productivity). Other than that, phrasal verbs and repeatedly used phrases or formulaic phrases were marked as one word by adding + in the middle shown in Table 4.1 above. Following Mehnert (1998), who suggested

that phrasal verbs or phrases should be counted as one linguistic item. This is because if they were to be separated, it would bring a different meaning than the one intended.

Additionally, in the lemmatisation process, unclear words (marked as xxx in the transcripts), pauses (…) and fillers (uh, em, er etc.) including retracings or reformulations and repetitions were excluded from the analysis (all measures of complexity and accuracy). These disfluencies were analysed for fluency measures. Furthermore, proper nouns, non-English words (Malay vocabulary) and acronyms (e.g. CEO, BMI, AM/PM etc.) were also removed as these words could falsely increase the number of tokens and types (Courtney, 2014) and thus would be an invalid index of students' language performance (Treffers-Daller et al., 2016). Nevertheless, the omission of words such as determiners and wrong usage of existing words, such as the preposition *in* where *on* is evidently to be used, were left uncorrected.

4.4 Measures of Analysis in Speaking

Various types of units were employed to assist in measuring the spoken data. Since spoken data is different from written forms, it is crucial for the speech to be segmented into reliable and accurate units. Previous studies employed the C-unit and T-unit but these units were found to be inadequate because they do not capture the intricacy and spontaneity of oral production (Foster et al., 2000).

The C-unit is defined as utterances that carry words, phrases, or sentences, grammatical and ungrammatical, which give referential or pragmatic meaning (Pica, Holliday, Lewis, & Morgenthaler, 1989). However, the C-unit does not grasp important characteristics of spoken data such as independent sub-clause unit, intonational or syntactic unit. (Foster et al., 2000). Another popular unit is the T-unit. Unlike the C-unit, the T-unit consists of the main clause together with any other dependent clauses or the shortest unit without leaving the sentence fragments as residue (Hunt, 1966). However, similar to the C-unit, it also has its shortcomings.

The T-unit is most appropriate for use with written discourse but is insufficient to give a full analysis of speech forms that include dysfluency, reformulations, hesitations and repetitions (Foster et al., 2000). Therefore, there is a need to find a standardized measure because the C- and T-units do not take account of speakers' elliptical construction, and thus do not cover all aspects of speaking resulting in the omission of valuable spoken data (Foster et al., 2000).

To solve this issue, the Analysis of Speech unit, known as the AS-unit, was introduced as the basic unit to capture the interactive data and enable segmentation of the participants' speech. The AS-unit would be most appropriate for analysing oral speech data because of its reliability and validity (Foster et al., 2000) and it has been adopted in several recent studies (Malicka & Levkina, 2012; Qian, 2014). The AS-unit consists of an independent clause or a sub-clausal unit with any associated dependent clause(s) within a single speaker utterance (Foster et al., 2000). In contrast with the C-unit and T-unit, Foster et al. (2000) further elaborated that:

> *"An independent clause will be minimally a clause including a finite verb… An independent sub-clausal unit will consist of: either one or more phrases which can be elaborated to a full clause by means of recovery of ellipted elements from the context of the discourse or situation… A subordinate clause will consist minimally of a finite or non-finite Verb element plus at least one other clause element (Subject, Object, Complement or Adverbial)."* (p. 365-366).

Therefore, in the current study, the AS-unit was employed as the main unit of analysis of the complexity, accuracy and fluency of the participants' speech.

4.4.1 Complexity

The speaking transcriptions were made into several sets of copies and saved separately in different folders from the original files. Two sets of copies were used for the analysis of

speaking complexity. The first copy was lemmatized to analyse lexical complexity and the second copy remain unchanged for the analysis of syntactic complexity.

4.4.1.1 Lexical Complexity

Numerous types of statistical measures cover different aspects of speaking complexity. Therefore, in order to answer the research question, several measures of complexity were used and are described in this section. In earlier studies, the type/token ratio (*TTR*) was used to capture lexical diversity in a given language sample (Treffers-Daller et al., 2016). *TTR* refers to the number of different words (types) divided by the overall number of words (Malvern & Richards, 2002). However, *TTR* is argued to be unreliable because of its sensitivity to the text length (Malvern & Richards, 2002). Simply put, a large number of tokens (i.e. longer text) will result in a lower *TTR* score. This can be resolved through the *D* measure, which can be automatically calculated within CLAN by using *VOCD* for each transcription file. *VOCD* calculates the *D* values through random sampling for 100 trials of 35-50 tokens to produce a curve of *TTR* against token size and find the best fit by altering the value of the *D* (McCarthy & Jarvis, 2007). This method is claimed to be a reliable measure of lexical diversity and has been used in previous studies (Malvern & Richards, 2002; McCarthy & Jarvis, 2007, 2010).

However, the value of *D* is more likely to be affected by text length in speech rather than in written text (Koizumi & In'nami, 2012). The token number produced by the students was uneven despite the fact that the time given for the speaking test was the same for every student in each test. The highest token was 589 and the lowest was 228. Therefore, the analysis of students' vocabulary could potentially been skewed; those who produced a larger sample of speech or text would potentially have had a greater opportunity to use a greater variety of words (Wright, Silverman, & Newhoff, 2003) than those who produced a smaller sample. In order to overcome this inherent weakness, the length of speech or written text should be controlled or kept within reasonable limits (Treffers-Daller, 2013). It is suggested that the score

of *D* can be only be calculated if a minimum of 50 tokens is achieved (Treffers-Daller, 2009) but is more reliable within a stable range of 100-400 tokens (McCarthy & Jarvis, 2007).

Since the minimum token produced was 228 in the present study, 200 tokens were truncated from the middle section of each transcription. This was done because different text length can affect the measure of lexical diversity, hence text length should be consistent and equal (Jarvis, 2002; Treffers-Daller et al., 2016). Taking out the middle section of the speech led to a more equitable analysis, making it possible to distinguish groups of different language skills (Koizumi & In'nami, 2012; Wright et al., 2003). It is arguable that not all speakers will perform their best in the middle of their speech and some might do well earlier or at the end of their conversation (Wright et al., 2003). Nevertheless, after careful consideration and thorough analysis of each transcription, the majority of the students repeated the questions in the first few sentences and also summarised the repeated information at the end of their speech. This agrees with Treffers-Daller et al. (2016) who stated that reproductions of lexical and syntactic structures occur most frequently at the beginning of the text. In other words, these were repetitive and unauthentic. Consequently, the authenticity and complexity of the language increased during the middle section where students were elaborating and explaining new ideas.

Besides *D,* another method to measure lexical diversity is by using Giraud's Index (G). The *G* measure is "simple transformation of *TTR* by taking a square root of a denominator and adjusting the values" (Koizumi, 2012, p. 65). However, similar to *D, G* is also sensitive to text length. Regardless of its drawback, van Hout and Vermeer (2007) stated that *G* is often a better transformation and provides the most satisfactory results. Furthermore, Daller (2010) argued that *G* has been demonstrated to be stable and robust. Tidball and Treffers-Daller (2007) found that both *D* and *G* proved to be a valid measure of lexical diversity, with a large and significant correlation with each other for their data. This pattern of findings was also found to be similar in Daller and Xue (2007), who concluded that both measures are considered to be the most

appropriate to be used in the context of lexical diversity. Therefore, in the current study, *D* and *G* were used to measure lexical diversity and both measures were deemed essential in order to compare results.

4.4.1.2 Lexical Productivity

Apart from lexical diversity, it was necessary to measure the lexical productivity (e.g. lexical content classes or lexical density) of the speaking transcriptions; noun, verb and adjective, in particular, to evaluate whether these lexical items or content words contributed the most to students' vocabulary performance. Numerous vocabulary studies count both content and function words all together but then again Bulte et al. (2008) argued that lexical productivity is used to assess the ability to use linguistic form units that have semantic-conceptual meaning. In other words, a varied use of content words indicates good knowledge of semantics and vocabulary; by contrast, function words only reflect students' grammatical competence. However, as there were fewer than 50 tokens within each word class, *D* could not be calculated. Therefore, all nouns, verbs and adjectives were counted as content words which added together, produced more than 50 tokens. *G* can be calculated in two different ways that have been applied in several studies to both verbs and adjectives (Bulte et al., 2008; Courtney, 2014; Treffers-Daller, 2009): 1) the noun types ratio divided by the square root of noun tokens (N types / √N tokens) and 2) the noun types ratio over the square root of all tokens (N types / √all tokens). The latter was used in this study, namely the content word types ratio over the square root of all tokens (content word types / √all tokens), because it is more desirable as the same denominator is used for all calculations (Treffers-Daller, 2009).

4.4.1.3 Syntactic Complexity

Measures of syntactic complexity focus on the use of complex and sophisticated structures of the language (Inoue, 2016). Previous studies mainly used a general variable such as length-

based and/or subordination-based variable as a measure of syntactic complexity. However, such methods might be insufficient to analyse syntactic complexity as a whole and could be unable to capture the multidimensional nature of the language development of L2 learners (Norris & Ortega, 2009). To overcome such weakness, Norris and Ortega (2009) suggested three measures which were considered suitable for this study because students' level of proficiency was categorised as within or above intermediate. These three measures were namely: 1) *length*, which is measured by an overall length-based metric (e.g. words per chosen unit); 2) *subordination*, which is measured by the total number of clauses; and 3) *sub-clausal*, which is measured by clause length-based metric. Therefore, in the current study, the measures were operationalized as: 1) mean length of AS-unit; 2) ratio of subordination; and, 3) mean length of clauses.

4.4.2 Accuracy

Ellis and Barkhuizen (2005) suggested that both global and local measures of accuracy tend to correlate closely. These measures refer to the percentage of error-free clauses (Skehan & Foster, 1999) and the number of errors per 100 words (Mehnert, 1998). The former refers to the number of errorless clauses divided by the total number of all clauses including independent, sub-clausal units or subordinate clauses, multiplied by 100. The former measure of accuracy has been shown to be more realistic and sensitive in identifying varieties of errors between experimental conditions (Skehan & Foster, 1999). Meanwhile, the latter refers to the number of errors divided by the number of tokens, multiplied by 100. Even though this measure is not sensitive in segmenting clauses like the former, Inoue (2016) claimed that counting errors per 100 words is more reliable since it does not involve clauses and unit-based measures (AS-unit, T-unit and C-unit) which can be problematic as they can be difficult to define. As the two measures look into different aspects of accuracy, the present study employed both

measures to offer a more robust analysis of the participants' speaking accuracy. However, such measures do not differentiate between the degree of severity of errors produced (Storch & Wigglesworth, 2007) which may cause difficulties in defining and determining what constitutes an error (Ellis & Barkhuizen, 2005). In order to solve this matter, Storch and Wigglesworth (2007) proposed that errors can be classified into three categories: 1) syntax (e.g. word order, word omission); 2) morphology (e.g. verb tense, subject-verb agreement, misused of determiners and preposition, errors in word form); and 3) lexis (e.g. incorrect word choice).

Similar to complexity, the unlemmatised version of the speaking transcriptions within the same 200-words range was used for accuracy. However, since each transcription was truncated from the middle section, errors emerging at the point of the 200-words cut were disregarded. Therefore, in this present study, error-free clause is defined as no errors in syntax, morphology or lexis. Nevertheless, phonological features such as intonation and stress were excluded from the analysis as these were beyond the scope of the study.

4.4.3 Fluency

In this study, fluency is described as smooth, uninterrupted and hesitation-free speech within a time constraint (Tavakoli, 2011). Following the definition, fluency can be divided into three main dimensions which according to Tavakoli and Skehan (2005) are: 1) breakdown fluency, to assess pauses and silence in speech; 2) repair fluency, to assess hesitation phenomena related to dysfluency; and 3) speed fluency, to assess the articulation of language produced. However, as mentioned in the previous chapter, intonation, stress and pronunciation were not part of the study, thus 3) was disregarded.

In measuring 1) breakdown fluency, the amount of pausing is a good predictor of perceptive fluency (Guz, 2015). Therefore, filled and unfilled pauses were taken into account as previous studies revealed that frequency of these pauses significantly differentiates between fluent and

dysfluent speakers (Lennon, 1999). Moreover, students who are more fluent spend less time pausing (Ellis & Barkhuizen, 2005). Filled pauses were referred to as non-lexical fillers or meaningless sounds such as *um, uh, er* etc. (Kang, Rubin, & Pickering, 2010). On the other hand, unfilled pauses can be identified as silent pauses equal to or more than 250 milliseconds (de Jong & Bosker, 2013). The length of the silent pauses was measured by listening to each transcription and examining the waveform using Goldwave software. The number of pauses that were equal or more than 250 milliseconds were counted and added to CLAN to be analysed. Besides that, it is arguable that pauses may occur between mid-clause and end-clause in a single utterance.

Meanwhile, 2) repair fluency in this study was analysed in four ways according to Skehan and Foster (1999), namely: a) false starts (e.g. abandoned utterances before completion); b) repetitions (e.g. verbatim repeated words, phrases or clauses without modification); c) reformulations (e.g. repeated phrases and clauses with some modification); and d) replacements (e.g. lexical items that immediately replenished for another).

To sum up, a full description of the 11 measures used in the current study is presented in Table 4.2 below.

Table 4. 2*: Measures of complexity, accuracy and fluency*

Dimensions	Measures	Abbrev.	Definition
Lexical Complexity	Lexical diversity (D)	LCD	Computed by CLAN using VOCD.
	Lexical diversity (G)	LCG	Type ratios divided by the square root of all tokens.
Lexical Productivity	Lexical density (D)	LPD	Computed by CLAN using VOCD +sm;*,|n +sm;*,|v +sm;*,|adj
	Lexical density (G)	LPG	The content word types ratio divided by the square root of all tokens.
Syntactic Complexity	Mean length of AS-unit	SMA	The total number of tokens divided by the number of AS-units.
	Ratio of subordination	SRS	The total number of clauses divided by the number of AS-units.
	Mean length of clauses	SMC	The total number of tokens divided by the number of clauses.
Accuracy	Percentage of error-free clauses	APE	The total number of errorless clauses divided by the total number of all clauses, multiplied by 100.
	Number of errors per 100 words	ANE	The total number of errors divided by the number of tokens, multiplied by 100.
Fluency	Number of pauses	FNP	The total number of pauses (filled and unfilled) divided by the total time of speech in seconds, multiplied by 60.
	Number of repairs	FNR	The total number of repairs (false starts, repetitions, reformulations and replacements) divided by the total time of speech in seconds, multiplied by 60.

4.5 Interrater Reliability

A total of six speaking data (three from each group) was recoded by an expert to check the reliability of the researcher coding for the measures of complexity, accuracy and fluency. Mackey and Gass (2005) suggested using a simple percentage that is "ratio of all coding agreements / total number of coding" (p. 243). A percentage above 75% is considered as good and acceptable (Mackey & Gass, 2005). In the first round, there were some disagreements and

thus reassessment was made for all coded measures by the researcher. In the second round, comparison was made and showed that both complexity, fluency scored 88% and 917 respectively. The remaining percentage of both measures were discussed between the researcher and the expert which included the coding, the types and tokens (complexity), and distinguish between reformulation and repetition (fluency). As for accuracy, however, the rating was low. Therefore, to resolve this issue, a native speaker of English expert recorded the accuracy measures for certain types of errors which involved mainly on the inappropriate native-like use. The third round for accuracy was 92% and the remaining percentage were discussed.

4.6 Quantitative Analysis of Speaking

This section presents a quantitative analysis of complexity, accuracy and fluency of speaking. The analysis will assess the changes made between groups and over time. The descriptive results will be presented first, followed by the inferential results.

4.6.1 Descriptive Results

The results of the procedures were first entered into an Excel spreadsheet and later imported into SPSS for statistical analysis. Descriptive statistics were generated to assess data distribution and to provide an overview score of speaking complexity, accuracy and fluency for both groups. Table 4.3 below shows the mean score (*M*) and standard deviation (*SD*) of students' speaking performance at pre and post-test. To reiterate, the pre-tests were collected at the beginning of the semester (February/March), followed by the intervention for eight weeks and the post-test then took place a week after (April/May).

Table 4. 3: *Comparison of results within 200 words cut*

Dimensions	Measures	EIST Group (*N*=30)				ST Group (*N*=29)			
		Pre-test		Post-test		Pre-test		Post-test	
		M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Lexical Complexity	LCD	41.20	7.90	57.35	14.25	52.66	9.63	49.07	14.49
	LCG	5.82	0.49	6.90	0.64	6.53	0.54	6.40	0.90
Lexical Productivity	LPD	37.70	12.89	60.31	25.39	49.69	15.98	51.60	8.74
	LPG	3.36	0.41	4.13	0.57	3.92	0.48	3.89	0.65
Syntactic Complexity	SMA	14.09	2.78	14.05	2.21	14.01	2.08	14.05	2.51
	SRS	2.03	0.38	1.94	0.33	1.92	0.31	2.03	0.40
	SMC	6.98	0.92	7.33	0.96	7.37	0.83	6.95	0.74
Accuracy	APE	68.72	12.72	72.15	14.22	66.56	12.53	68.79	11.99
	ANE	5.60	2.50	4.88	2.51	6.21	2.60	5.57	2.83
Fluency	FNP	12.93	5.05	10.14	4.39	11.94	4.37	7.08	4.59
	FNR	3.79	1.55	3.68	1.63	4.46	2.18	3.20	2.13

Note: M=Mean; SD=Standard Deviation; LCD=Lexical diversity (D); LCG=Lexical diversity (G); LPD=Lexical density (D); LPG=Lexical density (G); SMA=Mean length of AS-unit; SRS=Ratio of subordination; SMC=Mean length of clauses; APE=Percentage of error-free clauses; ANE=Number of errors per 100 words; FNP=Number of pauses; FNR= Number of repairs.

These results show that for the EIST group, overall speaking performance increased more than for the ST group. The mean scores of the EIST group at post-test were higher than the pre-test scores in both measures of *D* and *G* (for lexical complexity and productivity). Thus, this shows that students were able to produce a larger range of vocabulary and content words after the intervention was conducted. By contrast, the ST group students did not show any improvement for *D* or *G* (lexical complexity and productivity). As for accuracy, the mean score for APE increased in both groups over time which indicated that the percentage of error-free clauses increased. The mean score for ANE decreased from pre- to post-test which showed that students in both groups managed to reduce the number of errors per 100 words. The fluency scores (FNP and FNR) decreased which indicated that students in both groups produced fewer pauses and repairs during their post-test. However, for the measure of syntactic complexity,

the mean scores showed that both groups did make any drastic improvement. There was a slight increase in *SMA* and SRS for the ST group and *SMC* for the EIST group.

V 4.6.2 Inferential Statistics Results

Tests of normality were conducted and a histogram created for each test of every measure; these results were examined in order to select the appropriate data analysis method. The Kolmogorov-Smirnov and Shapiro-Wilk tests showed that for both pre- and post-test measures of LCG, LPG, SMC and APE were normally distributed, $p > .05$. Therefore, these measures were then explored by performing 2x2 repeated measures ANCOVA using SPSS Generalized Linear Model and entering the vocabulary levels test as a covariate. Another similar 2x2 and ANOVA was also conducted without covariate. The vocabulary levels test was introduced because vocabulary and the speaking test scores were significantly correlated. The correlation indicated that a higher level of vocabulary might contribute to an increase in the speaking test scores. Therefore, the vocabulary levels test was included as a covariate for variables where parametric tests were used to reduce the error that might arise from the non-equal levels of vocabulary. The presence of a covariate was considered crucial to reduce errors from unexplained variance and eliminate biases caused by unmeasured confounding variables (Fields, 2018). In addition, there were no outliers to be found except for the measure of SMC post-test which had a higher value of 9.53. A 2x2 repeated measures ANCOVA was conducted both with and without outlier and the results revealed both had the same outcome. Therefore, for SMC measure, the result with the outlier is reported.

On the other hand, normality tests at both pre- and post-test for measures of LCD, LPD, SMA and SRS were statistically significant ($p < .05$) which indicated that the data distribution was significantly non-normal with a positive skew. However, for ANE, FNP and FNR, all pre-tests were normal but post-test were not normally distributed. Therefore, non-parametric analyses

were selected to compare the differences in these measures between the EIST and ST groups at the pre and post-test. The analyses were performed separately for each measures using a two-independent samples test, the Mann-Whitney U test, to compare groups at pre-test and post-test. In all cases, 2-tailed statistics were reported. Then, the comparison between pre and post-test was made separately for each group using the Wilcoxon Signed-Rank test. As it is not possible to include a covariate in such tests, the analysis did not involve controlling for vocabulary knowledge.

In the findings also, effect sizes (*d* and *r*) were calculated to indicate the magnitude of change / difference in the results. For Cohen's *d*, the following benchmarks were used: between-group contrasts, small = .4, medium = .7 and large = 1.0; and pre-post/within-group contrasts, small = .6, medium = 1.0 and large = 1.4. For *r*, the following benchmarks were used: small = .25, medium = .40, and large = .60 (Plonsky & Oswald, 2014). Furthermore, where scores differed significantly at pre-test, pre/post-test gain scores were calculated and analysed using a Mann-Whitney U test.

4.6.2.1 Lexical Complexity

Lexical Diversity, *D* (LCD)

Between groups: Pre- and post-test

The Mann-Whitney U test suggested that there was a significant difference between the EIST group (*Mdn* = 40.27) and the ST group (*Mdn* = 52.79) at the pre-test, $U = 725.00$, $z = 4.40$, $p < .001$, $r = .57$. Likewise, the test indicated that the scores for LCD of the EIST group (*Mdn* = 55.43) and the ST group (*Mdn* = 44.21) differed at post-test, $U = 278.50$, $z = 2.37$, $p = .018$, $r = .31$. The effect sizes were medium at the pre-test but small at the post-test.

Over time: Pre- to post-test

A Wilcoxon Signed Rank test showed that the LCD score for the EIST group significantly increased between pre-test (*Mdn* = 40.27) and post-test (*Mdn* = 55.43), $z = 4.66$, $p < .001$, $r = .60$ with a large effect size. However, LCD scores for the ST group at pre-test (*Mdn* = 52.79) and post-test (*Mdn* = 44.21) were not significantly different, $z = 1.18$, $p = .239$, $r = .15$. These results were confirmed by a Mann Whitney-U test on the pre to post-test gain scores, showing significantly higher gains for the EIST group (mean rank = 39.93) compared with the ST group, with a medium effect size (mean rank = 19.72) ($U = 137.00$, $z = 4.5$, $p < .001$, $r = .59$).

Lexical Diversity, *G* (LCG)

A 2x2 repeated measures ANCOVA with one between-subjects variable (group: EIST, ST) and one within-subjects variable (time: pre-test, post-test) and vocabulary test score as a covariate was conducted. This revealed no main effect of time, $F(1,56) = .91$, $p = .344$, $\eta_p^2 = .016$, or group, $F(1,56) = .64$, $p = .427$, $\eta_p^2 = .011$. However, there was a significant time*group interaction $F(1,56) = 29.42$, $p < .001$, $\eta_p^2 = .344$, as shown in Figure 4.1 below. A follow-up post-hoc analysis with Bonferroni correction revealed that the LCG score in the EIST group showed a statistically significant increase from the pre-test to post-test (*M* = -1.08, *95% CI* [-1.39, -.76], $p < .001$) . However, this was not the case for the ST group (*M* = .13, *95% CI* [-.18, .45], $p = .404$).

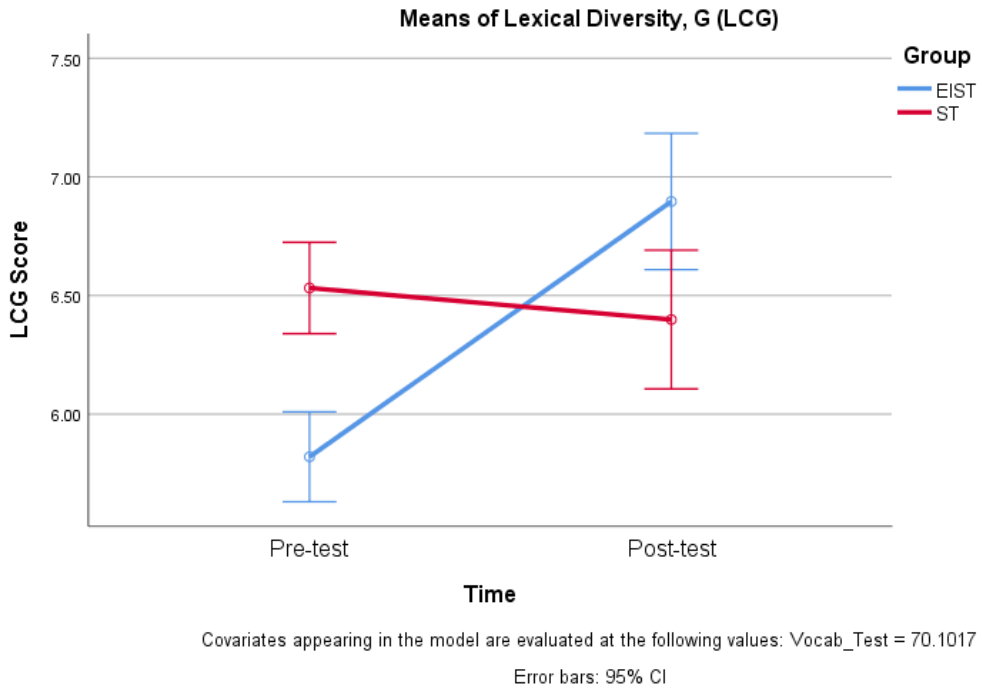

Figure 4. 1: Time*Group interaction for lexical diversity, G *(LCG)*

4.6.2.2 Lexical Productivity

Lexical Density, *D* (LPD)

Between groups: Pre- and post-test

The Mann-Whitney U test suggested that there was a significant difference between EIST group (*Mdn* = 35.55) and the ST group (*Mdn* = 51.79) at the pre-test, $U = 632.00$, $z = 2.99$, $p = .003$, $r = .39$, with a small effect size. However, the test indicated that the scores for LPD of the EIST group (*Mdn* = 57.38) and the ST group (*Mdn* = 47.59) did not differ at post-test, $U = 340.00$, $z = 1.44$, $p = .150$, $r = .19$.

Over time: Pre- to post-test

A Wilcoxon Signed Rank test showed that the LPD score on the EIST group significantly increased between pre-test (*Mdn* = 35.55) and post-test (*Mdn* = 57.38), $z = 4.06$, $p < .001$, $r = .52$, with a medium effect size. However, the LPD scores for the ST group at pre-test (*Mdn* = 51.79) and post-test (*Mdn* = 47.59) were not significantly different, $z = .05$, $p = .957$, $r = .07$.

The pre to post-test gain scores of LPD for the EIST group (mean rank = 36.73) were significantly higher than the ST group, with a medium effect size (mean rank = 23.03) (U = 233.00, z = -3.06, p = .002, r = -.40).

Lexical Density, *G* (LPG)

A 2x2 repeated measures ANCOVA with one between-subjects variable (group: EIST, ST), one within-subjects variable (time: pre-test, post-test) and vocabulary test score as a covariate was conducted. This revealed no main effect of time, $F(1,56) = .27$, $p = .608$, $\eta_p^2 = .005$, or group $F(1,56) = 2.58$, $p = .114$, $\eta_p^2 = .044$. However, there was a significant time*group interaction $F(1,56) = 18.98$, $p < .001$, $\eta_p^2 = .253$ shown in Figure 4.2 below. A follow-up post-hoc analysis with Bonferroni correction revealed that the LPG score in the EIST group showed a statistically significant increase from the pre-test to post-test (M = -.77, *95% CI* [-1.02, -.51], $p < .001$). However, there was no significant difference for the ST group (M = -.02, *95% CI* [-.23, .28], p = .855).

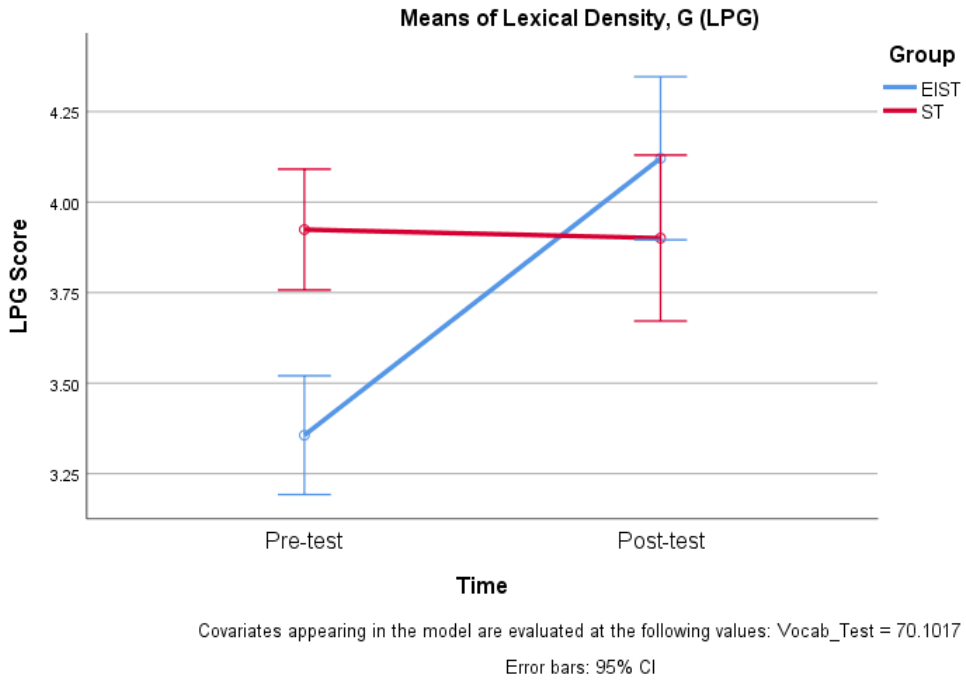

Figure 4. 2: Time*Group interaction for lexical density, G (LPG)

4.6.2.3 Syntactic Complexity

Mean Length of AS-unit (SMA)

Between groups: Pre- and post-test

The Mann-Whitney U test suggested that there was no significant difference between the EIST group (*Mdn* = 13.33) and the ST group (*Mdn* = 13.33) at the pre-test, $U = 485.50$, $z = .36$, $p = .719$, $r = .05$. Likewise, the test indicated that the scores for SMA of the EIST group (*Mdn* = 13.81) and the ST group (*Mdn* = 13.33) did not differ at post-test, $U = 441.50$, $z = .10$, $p = .921$, $r = .01$.

Over time: Pre- to post-test

A Wilcoxon Signed Rank test for SMA showed no difference between pre-test (*Mdn* = 13.33) and post-test (*Mdn* = 13.81), $z = .27$, $p = .791$, $r = .03$ for the EIST group. Likewise, the SMA scores for the ST group at pre-test (*Mdn* = 13.33) and post-test (*Mdn* = 13.33) were not significantly different, $z = 1.27$, $p = .899$, $r = .17$.

Ratio of Subordination (SRS)

Between groups: Pre- and post-test

The Mann-Whitney U test suggested that there was no significant difference between the EIST group (*Mdn* = 2.00) and the ST group (*Mdn* = 1.85) at the pre-test, $U = 348.50$, $z = 1.31$, $p = .189$, $r = .17$. Likewise, the test indicated that the scores for SRS of the EIST group (*Mdn* = 1.88) and the ST group (*Mdn* = 2.00) did not differ at post-test, $U = 514.00$, $z = 1.20$, $p = .230$, $r = .16$, in relation to SRS.

Over time: Pre- to post-test

For the Wilcoxon Signed Rank test, there was no difference in the SRS score for the EIST group between pre-test (*Mdn* = 2.00) and post-test (*Mdn* = 1.88), $z = 1.13$, $p = .258$, $r = .15$.

Similarly, SRS scores for the ST group at pre-test (*Mdn* = 1.85) and post-test (*Mdn* = 2.00) were not significantly different, z = .93, p = .350, r = .12.

Mean Length of Clauses (SMC)

A 2x2 repeated measures ANCOVA with one between-subjects variable (group: EIST, ST), one within-subjects variable (time: pre-test, post-test) and vocabulary test score as a covariate was conducted. This revealed no main effect of time, $F(1,56)$ = .49, p = .488, η_p^2 = .009, or group $F(1,56)$ = .01, p = .933, η_p^2 < .001. However, there was a significant time*group interaction, as shown in Figure 4.3 below. $F(1,56)$ = 7.50, p = .008, η_p^2 = .118. A follow-up post-hoc analysis with Bonferroni correction revealed that the SMC score in the EIST group showed no significant difference from the pre to post-test (*M* = -.37, *95% CI* [-.78, .04], p = .078). However, there was a statistically significant decrease for the ST group (*M* = -.44, *95% CI* [.02, .85], p = .041).

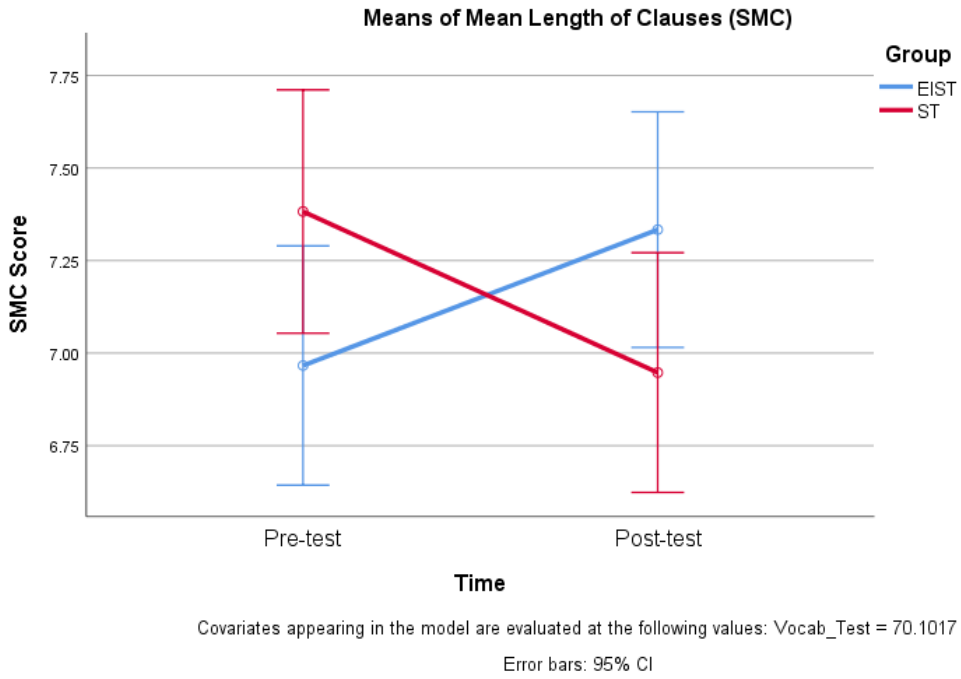

Figure 4. 3: Time*Group Interaction for mean length of clauses (SMC)

4.6.2.4 Accuracy

Percentage of Error-free Clauses (APE)

A 2x2 repeated measures ANCOVA with one between-subjects variable (group: EIST, ST), one within-subjects variable (time: pre-test, post-test) and vocabulary test score as a covariate was conducted. This revealed no main effect of time, $F(1,56) = .55$, $p = .460$, $\eta_p^2 = .010$, or group $F(1,56) = 2.54$, $p = .116$, $\eta_p^2 = .043$. Likewise, there was no significant interaction between time*group $F(1,56) = .05$, $p = .819$, $\eta_p^2 = .001$, as shown in Figure 4.4 below.

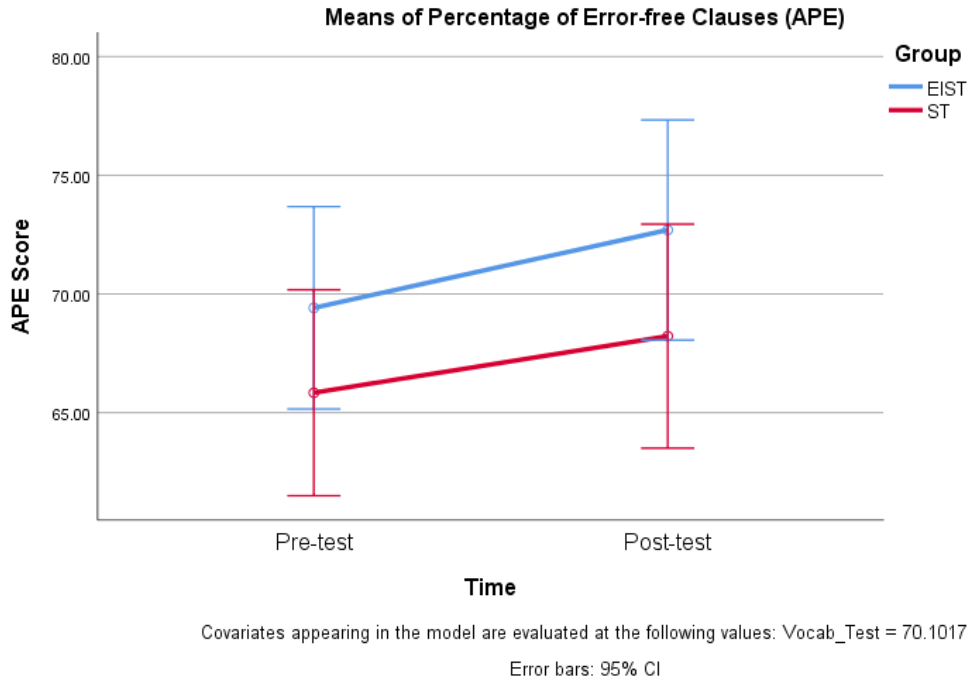

Figure 4. 4: Means of percentage of error-free clauses (APE) between both groups

Number of Errors per 100 Words (ANE)

Between groups: Pre- and post-test

The Mann-Whitney U test suggested that there was no significant difference between the EIST group (*Mdn* = 5.25) and the ST group (*Mdn* = 4.50) at the pre-test in relation to ANE, $U = 482.50$, $z = .72$, $p = .470$, $r = .09$. Likewise, the test indicated that the scores for ANE of the EIST group (*Mdn* = 5.50) and the ST group (*Mdn* = 5.00) did not differ at post-test, $U = 474.00$, $z = .59$, $p = .553$, $r = .08$.

Over time: Pre- to post-test

For the Wilcoxon Signed Rank test, there was no significant difference for the EIST group in ANE score between pre-test (*Mdn* = 5.25) and post-test (*Mdn* = 5.50), *z* = 1.40, *p* =.160, *r* = .18. Similarly, the ANE scores for the ST group at pre-test (*Mdn* = 4.50) and post-test (*Mdn* = 5.00) were not significantly different, *z* = 1.01, *p* = .312, *r* = .13.

4.6.2.5 Fluency

Number of Pauses (FNP)

Between groups: Pre- and post-test

For FNP, the Mann-Whitney U test suggested that there was no significant difference between the EIST group (*Mdn* = 13.02) and the ST group (*Mdn* = 11.67) at the pre-test, *U* = 493.00, *z* = .879, *p* = .379, *r* = .11. However, the test indicated that the scores for FNP of the EIST group (*Mdn* = 10.09) and the ST group (*Mdn* = 5.58) differed at post-test, *U* = 596.00, *z* = 2.44, *p* = .015, *r* = .32, with a small effect size.

Over time: Pre- to post-test

For the Wilcoxon Signed Rank test, the FNP score of the EIST group between pre-test (*Mdn* = 13.02) and post-test (*Mdn* = 10.09) was significantly different, *z* = 2.80, *p* = .005, *r* = .37, with a small effect size. Likewise, there was a statistically significant decrease in the FNP score for the ST group between pre-test (*Mdn* = 11.67) and post-test (*Mdn* = 5.58), *z* = 3.96, *p* < .001, *r* = .51, with a medium effect size.

Number of Repairs (FNR)

Between groups: Pre- and post-test

The Mann-Whitney U test suggested that there is no significant difference between the EIST group (*Mdn* = 3.62) and the ST (*Mdn* = 4.51) group at the pre-test, *U* = 354.50, *z* = 1.22, *p* = .222, *r* = .16. Likewise, the test indicated that the scores of FNR for the EIST group (*Mdn* =

3.75) and the ST group (*Mdn* = 2.73) did not differ at post-test, $U = 539.50$, $z = 1.59$, $p = .113$, $r = .21$.

Over time: Pre to post-test

For the Wilcoxon Signed Rank test, the FNR scores for the EIST group at pre-test (*Mdn* = 3.62) and post-test (*Mdn* = 3.75) were not significantly different, $z = .29$, $p = .770$, $r = .04$. However, the FNR scores for the ST group significantly decreased between pre-test (*Mdn* = 4.51) and post-test (*Mdn* = 2.73), $z = 2.79$, $p = .005$, $r = .36$, with a small effect size.

4.7 Qualitative Analysis of Speaking

The speaking data were next examined with the aim of gaining insights into students' speech discourse in terms of the kind of language they were using at both pre- and post-test and what changes occurred between the two time points. Areas of interest were vocabulary (lexical complexity), forms (syntactic complexity), errors (accuracy), and disfluency (fluency) produced at both pre- and post-test. The analysis was considered to be qualitative in so far as judgements were made about levels of complexity, types of errors, and so forth. However, there was also an element of quantification in order to establish patterns with respect to such language use. Such a combination of qualitative and quantitative analysis is a recognised analytical approach, particularly in mixed-methods research (Sandelowski, 2001).

4.7.1 Lexical Complexity and Productivity

For this section, the measures of lexical complexity and productivity were combined in one section as both measures assessed students' lexical knowledge and development over time. According to Bybee (2007), lexical frequency is relevant in the study of language and can be calculated into two ways; token frequency and type frequency. Token frequency is the frequency of occurrence of a unit, either a word of a morpheme, whereas type frequency refers

to the patterns of the language (e.g. lexical items, morphological and syntactic patterns etc.) (Bybee, 2007).

The tokens with the highest frequency in both groups were found to be mostly of function words especially pronouns (*we*), articles (*a, the*), prepositions (*to*), conjunctions (*and*), modal verbs (*can*) and auxiliary verbs (*have).* However, even though students produced more functions words, Bybee (2007) argued that they do not contribute to the overall language proficiency. Therefore, it is crucial to count the lexical items as it is an important determinant of productivity. As mentioned in the previous chapter, lexical items in this study referred to the production of content words, primarily nouns, verbs and adjectives. Students in both groups produced nouns (tokens) the most in both tests followed by verbs and the least produced were adjectives.

Bybee (2007) suggested calculating type frequency which refers to the number of distinct lexical items within the construction of the language. As mentioned in the quantitative section of the analysis, it was found that the EIST group made progress in both measures of lexical complexity and lexical productivity. However, this was not thoroughly explored, thus qualitative analysis was needed to examine what types of lexical items were produced in both groups in two different time points. The frequency of nouns increased from pre- to post-test for the EIST group, whereas the ST group's noun production decreased. The noun types produced by the students in both groups can be classified into the following different content categories: sports, education, people and lifestyles, leisure and pastimes, and health. These categories, found in both tests, were expected as they were linked to the thematic areas that had recently been covered within the intervention (see Section 3.5.1 for the topics and themes taught during the intervention). Certain nouns appeared frequently and were produced by the majority of the students in the pre-test but did not appear at all during the post-test such as *attitude* and *happiness, class, junk+food* and *tiredness*. Although these words were primarily taught during

the intervention and produced by the majority of students, they were not retained and available for use at the end of the semester. By contrast, some noun types were used in both pre- and post-tests such as *diet, diabetes, health* and *stress, activity, health* and *lecturer*. Meanwhile, the frequency of other nouns dramatically increased over time and varied between groups, for example, *body, food, meal, student* and *time* (EIST group); and *life, lifestyle, people* and *thing* (ST group). Although the frequency of nouns produced by the EIST group was higher than the ST group, concrete nouns were found to be produced the most often over time. However, abstract nouns (e.g. *freedom, expectation, quality, happiness, attitude. success, tiredness*) were more prominent in the ST group even though the frequency of nouns was lower than for the EIST group.

Meanwhile, the frequency of verbs in both groups increased from pre- to post-test. The verb production reflected the subject matter of the questions in the speaking tests which the students were asked based on their personal experience. Students drew on relevant active verbs relating to the topic of the questions. For example, both groups were found to be using very similar types of dynamic verbs (e.g. *drink, eat, sleep, teach, travel*) to describe actions of the subject. It is also crucial to note that, in the post-test, the ST group were found to be more revealing and expressive in conveying their ideas with the production of stative verbs (e.g. *I know, I think, believe, I prefer*), formulaic phrases (e.g., *in+my+opinion, on+the+other+hand, and+so+on*) and repeated usage of discourse markers (e.g., *okay, right, well*).

As for adjectives, the EIST group produced a low frequency of adjectives over time, whereas the production of adjectives for the ST group increased from pre- to post-test. The majority of the adjectives produced in both tests were demonstrative adjectives and these were produced consistently in both the pre- and post-tests. Thus, overall the participants produced a high frequency of adjective tokens and types. Adjectives such as *healthy* and *important* were found

to be frequently uttered by the majority of the students in both groups and consistently stable throughout both tests. Some adjectives were used quite frequently in the pre-test: *positive, confident, comfortable* (EIST group); and *sleepy, lazy* (ST group). Moreover, students in both groups, particularly in the EIST group, were more often found to create new lexemes using adjective-forming suffixes in the post-test compared to the pre-test such as *energetic, enjoyable, suitable, successful* and *productive.*

4.7.2 Syntactic Complexity

To reiterate, the quantitative results showed that students in both groups did not make any improvement on any of the measures of syntactic complexity. As mentioned in Section 4.4.1.1, analysis of the speaking data focussed on an excerpt of 200 words (i.e. tokens). From the quantitative analysis, in the pre-test, the speaking data for the EIST group showed that students produced within the range of 9 to 19 AS-units and 22 to 38 clauses, whereas for the ST group, the AS-units were within the range of 11 to 19 and clauses were within 22 to 36. Meanwhile, in the post-test, students in both groups produced approximately the same range of AS-units and clauses. For the EIST group, the range of AS-units and clauses was within 11 to 19 and 21 to 36, respectively, whereas for the ST group, students produced 10 to 23 AS-units and 24 to 36 clauses.

Therefore, a thorough qualitative analysis was conducted to explore the types of sentence form in the students' speech data. The results showed that students in both groups used simple sentences (subject + verb) quite frequently when introducing a new idea. As students started to elaborate their ideas, the usage of compound (main clause + coordinating conjunction + main clause) and complex sentences (main clause + subordinating conjunction + dependent clause) emerged accordingly throughout their speech data. These sentence forms were prominent at both pre- and post-test. In compound and complex sentences, coordination conjunctions such

and, *but* and *so;* and subordinating conjunction such as *because* were found frequently within one AS-unit. A few examples are shown below to illustrate how the speaking data were coded, following Foster et al. (2000) (see Section 4.4). Grammar errors and disfluency were disregarded in this analysis of syntactic complexity. The coding symbols used in this analysis were: **|** = the end of AS-unit, and **::** = the end of clauses.

Pre-test We must communicate with others **|** We need more friends**::** because friends will give us more happiness **|** We can't (cannot) live without friends**::** because we need others **|**

Total AS-units: 3 **Total clauses: 5**

(EIST, Student 02)

A good relationship with other**::** can make you easily communicate with other **|** You also can form the best teamwork for your assignment**::** and can acquire good result **|**

Total AS-units: 2 **Total clauses: 4**

(ST, Student 18)

Post-test If we practice a bad eating habit**::** for example eating too much food that contain fats**::** will cause us sleepy and lazy **|** It will also lead to our body**::** to feel weak and tired **|**

Total AS-units: 2 **Total clauses: 5**

(EIST, Student 02)

Carbohydrate such as rice can give energy to our body**::** to do work **|** Eating a lot of carbohydrate are not good**::** because it change to fat**::** and we get tired easily**::** and you can become obesity **|**

Total AS-units: 2 **Total clauses: 6**

(ST, Student 18)

4.7.3 Accuracy

The quantitative results showed that students produced more errored clauses in both tests. The production of errors was also high and repeated within those errored clauses. In other words,

within one errored clause, error could occur more than once. Errors that emerged in the speaking data were frequent in both groups, thus contributed to the poor performance in all statistical measures of accuracy (see Section 4.6.2.4). Students' success at applying the grammatical items being taught was analysed and reported, including plural and singular form, subject-verb agreement, determiners, and word form (part of speech). These errors were sorted into three main categories: syntax, morphology, and lexis.

4.7.3.1 Syntactical Error

Errors in syntax were mostly related to word omission. These errors were deletion of grammatical items such as auxiliary verbs, determiners, prepositions etc. The results showed that the ST group was found to ignore more grammatical items than the EIST group in both pre- and post-test and thus the language produced was grammatically incorrect, even though it did not impede the intended meaning. In some cases, omission which is labelled as **X** in the example below involved lexical items but these were not as frequent as the omission of grammatical items.

Pre-test Besides that, we also need **X** take care of our health… **(to)**

(EIST, Student 18)

…eating a balance **X** to get health… to ensure our body **X** healthy **(diet/meal, is)**

(ST, Student 15)

Post-test If we **X** not get enough sleep… **(do)**

(EIST, Student 18)

…this **X** a important thing to have… **(is)**

(ST, Student 15)

4.7.3.2 Morphology

Errors in morphology were the most frequent in both groups for both time points, compared to syntactic errors and errors in lexical choice. The most prominent errors were mostly related to subject-verb agreement and singular/plural from, determiners and word form. In this sub-section, the most frequent errors are presented first. Students were inconsistent with the usage

of plurality and subject-verb agreement. In English, the subject and verb must agree with one another in number, but this was not the case in the students' speech production. Besides that, in English singular nouns are made plural by adding *–s, -es* or *–ies* at the end depending on what letter a noun ends in. This was difficult to detect in spoken form compared to written form where it is visible. The audio of both tests indicated that many students omitted the –s sound to indicate a plural noun or a third person verb, as well as incorrectly overusing both features. The examples are as follows:

Pre-test If we ***doesn't*** get happiness in our life… The most important factor that ***help*** a person attain happiness…

(EIST, Student 17)

…because they did not have enough time and that ***make*** them look so tired and sleepy

(ST, Student 29)

Post-test The next one is have a healthy eating ***habits***… We will feel tired and ***doesn't*** have time to study.

(EIST, Student 17)

You can do all of these ***activity*** once a week or two ***time*** a week.

(ST, Student 29)

Moreover, the misuse of determiners such as the indefinite article and demonstrative pronouns was also visible. The incorrect use of an indefinite article or singular demonstrative before a plural noun was in evidence, especially in the EIST group, for example:

Pre-test …because sometimes when ***a*** problems come to us…

(EIST, Student 03)

Second, is having ***a*** healthy eating habits.

(ST, Student 21)

Post-test …and we can do ***this*** things while we at that time.

(EIST, Student 03)

For me why ***this*** factors help us to attain happiness in life… Okay, in my opinion, have ***a*** good relationships because…

(ST, Student 21)

The least frequent errors in morphology compared to the above, but still notable in the speaking data, were errors in word form. Errors in word form occurred when the incorrect part of speech was used. A word can have several forms, depending on whether it is being used as a noun, verb, adjective or adverb. However, from the results, most students had difficulties in differentiating between noun and adjective, for example:

> …being *obesity* will make us be more easily tired.
>
> (Post-test: EIST, Student 04)

> …because healthy food can help their thinking better and can be good to their *healthy*.
>
> (Post-test: EIST, Student 28)

> I think junk food is not really good because it can effect on their *healthy*.
>
> (Pre-test: ST, Student 06)

4.7.3.3 Errors in Lexical Choice

Errors in this category were related to poor choice of vocabulary used by the students, e.g., vocabulary items which were semantically incorrect and may impede the intended meaning of the spoken utterance. These were evident at both time points for both groups. Moreover, non-existing words were also created by the students in the speaking tests which was only found in the pre-test. However, these non-existing words did not emerge in the post-test speech data of both groups. Although, the message being delivered may have been understandable, these were still considered to be incorrect lexical choices. The examples are as follows:

> …and *the worsely* they can be sick.
>
> (Pre-test: EIST, Student 14)

> So with light exercise we do *oftenly* weekdays and heavy exercise we do every weekend.
>
> (Pre-test: ST, Student 04)

…and can ***provide*** you from illness.

(Pre-test: ST, Student 26)

…try to learn how to forgive others and forget their ***false***.

(Post-test: ST, Student 03)

4.7.4 Fluency

Based on the quantitative results of fluency (see Section 4.6.2.5), students in both groups managed to improve speaking fluency, especially in relation to breakdown fluency (i.e. reducing number of pauses) from pre- to post-test. However, there was no significant difference in repair fluency for the EIST group between pre- and post-test. Through qualitative analysis, the speaking fluency data were analysed in terms of how pauses and repair were used and where they occurred most often. What differentiated the two groups was the frequency with which disfluencies occurred.

4.7.4.1 Breakdown Fluency

Filled Pauses

Students in both groups tended to make more filled pauses in both the pre- and post-tests in contrast to unfilled pauses. Overall, the frequency of pauses (filled and unfilled) produced by the EIST group was higher than for the ST group in both tests. In the qualitative analysis, it was found that students in both groups tended to produce filled pauses when they were trying to explain or elaborate their ideas and connect them with the correct words, and after they produced repairs. In CLAN, filled pauses were coded as &-, for example:

Pre-test …<when we have> [//] *&-uh* when you have tired of studying *&-uh* you can play some games on the phone or laptop or you can play monopoly.

(EIST, Student 13)

If we *&-ah* sleep *&-ah* less than six hours maybe we can get sick or like a headache or something *&-ah &-ah* and I feel like a tired.

(ST, Student 07)

Post-test Healthy lifestyle *&-uh* is important *&-um* because *&-uh* when we sick we cannot do *&-uh* activity [//] *&-uh* more activity.

(EIST, Student 13)

Healthy eating habit is *&-ah* can help our body to stay healthy that we need our body *&-ah* to stay healthy because *&-ah* every day we must do a lot of work.

(ST, Student 07)

Unfilled Pauses

Unfilled pauses or silent pauses were found to be more prominent usually at the beginning of the sentence before a new idea was explained or elaborated. Unfilled pauses also occurred mid-clause but not as frequently as filled pauses. It is worth noting that the study did not look into pause length (e.g. short, medium or long pauses) but instead frequency of the unfilled pauses (i.e. how many repetitions of unfilled pauses) which occurred equal to or more than 250 milliseconds (see Section 4.4.2). For example:

Pre-test …then you have a happy life. (.) <the second> [/] (.) the second factor is *&-uh* have a good relationship with others.

(EIST, Student 15)

…from the best teamwork. (.) Next you also must to have a healthy lifestyle.

(ST, Student 09)

Post-test … and student must follow [/] follow the timetable. (.) So (.) that they get enough time to *&-uh* more [//] to do more work or task in a day (.) in a systematic way and avoid tiredness.

(EIST, Student 15)

So (.) eating [//] (.) student should eating properly for example eat some *&-uh* vegetable protein to help us required enough energy for the daily activities.

(ST, Student 09)

4.7.4.2 Repair Fluency

Types of repair fluency were sorted into repetition, reformulation, and replacement. False start is not shown because it was not present in the speaking data. Students were able to start each new sentence smoothly without abandonment or interruption.

Repetition

The results revealed that repetition was the most frequent out of all the repairs that occurred in the speaking tests. Students in both groups produced approximately the same frequency of repetitions in the post-test. However, there was a larger decrease in the number of repetitions by the ST group between pre- and post-test. Students in both groups were found to be repeating the same words or phrases as they were elaborating and explaining their ideas, which can be seen as another form of filled pauses. Repetitions in CLAN were coded as <original word> **[/]** followed by the repeated word, for example:

Pre-test *We* **[/]** *we* **[/]** <*we do not*> **[/]** *we do not* do something that will make they angry or (.) something.
(EIST, Student 01)

So <*we need to balance*> **[/]** *&-ah we need to balance* our *&-ah time* **[/]** t*ime &-ah* to (.) study.
(ST, Student 03)

Post-test Other than that <*there are*> **[/]** *&-uh there are* other things that we can do to overcome the tiredness.
(EIST, Student 01)

We can *aiming* **[/]** *aiming* something in *&-ah* our life.
(ST, Student 03)

Reformulation

In the quantitative analysis (see Section 4.6.2.5), both groups made progress between pre- and post-test but the ST group produced fewer reformulations in post-test than the EIST group and made a significant decrease. Through qualitative analysis, students were found to use reformulations when they repeated phrases or clauses but with some alteration mainly because

of incorrect grammar use and/or by adding new words they were left out in the previous sentence. In CLAN, replacements were coded as <incorrect word/phrase> **[//]** <corrected word/phrase>, for example:

Pre-test If you want to attain your happiness in life <*you must be*> **[//]** (.) *you must have* a positive thinking.

(EIST, Student 02)

<*they just*> **[//]** *if they don't have class they just &-ah &-ah* sleep at their room.

(ST, Student 06)

Post-test So after doing a lot of activities <we need to> **[//]** ↫ᔕ↫ <some rest to> **[//]** we need to rest *&-uh* and sleep according to the right time.

(EIST, Student 02)

(.) &-ah Other than that we also can pick exercise &-ah *we* **[//]** *that we* can enjoy.

(ST, Student 06)

Replacement

The frequency of replacement showed that students in the ST group produced more replacements that the EIST group. The qualitative analysis revealed that replacement used by the students in both groups was mainly to replace an incorrect word with another to suit the context better due to poor word choice being uttered in the previous sentence. Replacements were coded similarly to reformulations, for example:

Pre-test Next *&-uh* what the other contributing factor for a person to *attend* **[//]** *attain* happiness in our life?

(EIST, Student 12)

Have a healthy lifestyle is *&-ah* something about get enough sleep so that our life can be *&-ah* more *energy* **[//]** *energetic.*

(ST, Student 04)

Post-test So eating *health* **[//]** *healthily* can help you gain your energy and reduce your tiredness.

(EIST, Student 12)

> For point number two is have a *health* **[//]** *healthy* eating habits.
>
> (ST, Student 04)

4.8 Summary

This chapter discussed the results from both the quantitative and qualitative analysis of students' performance on the speaking pre- and post-tests. The quantitative data were presented in descriptive (Section 4.6) and inferential statistics (Section 4.7). The statistical results showed that the EIST groups produced a larger variety of word types and tokens compare to the ST group. However, the qualitative analysis revealed that the ST group produced more complex and sophisticated language (e.g. the presence of abstract words). Neither group made progress in syntactic complexity and accuracy. The results also revealed that the ST group was more fluent than the EIST group at post-test, with a decrease in the occurrence of disfluencies over time. The following chapter discusses the findings of the questionnaire and the interview investigating students' motivation towards learning and speaking English.

CHAPTER FIVE: ANALYSES & FINDINGS OF L2 MOTIVATION

5.1 Quantitative Analysis of Motivation

This chapter focuses on students' L2 motivation. The analyses of the data were carried out to answer one of the main research questions and investigate the impact of the intervention on motivation. To reiterate, the question is:

> 1(b). What are the effects of two different forms of language teaching (TBLT and TSLT) on students' motivation towards learning and speaking English?

For this variable, motivation was analysed using two methods: 1) quantitative, statistical analysis via SPSS with the questionnaire results presented first in this chapter, and followed by 2) qualitative analysis, using thematic analysis and presented descriptively.

5.2 Quantitative Analysis of Motivation

The item scores obtained from the questionnaire were entered into SPSS version 24 for statistical analysis. They were entered as 1 to 4, corresponding to the four-point Likert scale (1 – strongly disagree, 2 – disagree, 3 – agree, and 4 – strongly agree). In the process of data entry, 17 items had to be recoded (1=4, 2=3, 3=2 and 4=1) because they were negatively-worded. In addition, no missing data were found during the data entry. Overall, a higher score indicates more positive or higher levels of motivation.

The questionnaire dealt with L2 motivation towards the English language and students' speaking motivation. All of the 45 items were grouped into five sub-scales representing the main elements of the construct of motivation, namely; attitudes towards speaking in a second language, integrativeness, instrumentality, linguistic self-confidence and learning situation.

According to Dörnyei (1994) motivation is a multifaceted variable because it is considered as an abstract concept which cannot be easily observed and measured. Therefore, it was important to examine the internal consistency and validity of the questionnaire by calculating the Cronbach's Alpha reliability coefficients as presented in Table 5.1 below. These coefficients indicate how well these items represent the underlying construct and how well these items fit together within the sub-scales. The reliability coefficient is acceptable when $\alpha > 0.7$ (Gay et al., 2011).

Table 5. 1*: The reliability tests score*

Sub-scales	**Items**	**Pre-test**	**Post-test**
Attitude	22, 24, 25, 34, 37, 40, 43	.69	.67
Integrative	9, 26, 30, 33	.66	.65
Instrumental	2, 5, 12, 13, 17, 19, 41	.74	.65
Linguistic Self-confidence	8, 10. 14. 15, 16, 20, 21, 23, 27, 28, 29, 31, 32, 35, 36, 38, 42, 44, 45	.81	.75
Learning Situation	1, 3, 4, 6, 7, 11, 18, 39	.65	.67
	All items	.90	.83

The overall score in both tests showed that $\alpha > 0.7$ but within the subscales, some were slightly below 0.7. However, these scores were considered acceptable because they contained fewer items (Cziser & Kormos, 2009). On reflection, it would have been better to add more items for the sub-scales which had fewer items to increase the reliability. However, after several considerations, the questionnaire should not be too long due to time constraints. On the other hand, the interview analyses were also conducted which served to add to the reliability of the quantitative data.

5.2.1 Descriptive Results

A variety of statistical analyses were utilised to investigate the effect of two forms of TBLT on students' motivation towards learning and speaking English. Similar to the speaking test, the means for each student were averaged to calculate the overall motivation scales for both pre- and post-tests, as presented in Table 5.2 below.

Table 5. 2: *Descriptive statistics for the EIST and ST group over time*

Sub-scales	**EIST Group (*N*=30)**				**ST Group (*N*=29)**			
	Pre-test		**Post-test**		**Pre-test**		**Post-test**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Attitude	2.90	0.42	2.88	0.39	2.79	0.50	2.78	0.50
Integrativeness	2.91	0.54	3.00	0.41	3.17	0.38	3.13	0.45
Instrumentality	3.10	0.37	3.24	0.27	3.31	0.25	3.26	0.26
Linguistic Self-confidence	2.65	0.35	2.86	0.21	2.72	0.34	2.80	0.30
Learning Situation	2.75	0.42	3.05	0.21	2.79	0.33	2.84	0.36

Note: M=Mean; SD=Standard Deviation

From the results, it can be seen that the EIST group was more positively motivated after the intervention with the increase of the mean score in all aspects from pre- to post-test except for attitude. Meanwhile, for the ST group, they made a slight improvement in linguistic self-confidence and learning situation from pre- to post-test. However, they were less motivated in terms of attitude, integrative and instrumental with a slight decrease in the mean score after the intervention.

5.2.2 Inferential Statistics Results

Normality tests were conducted on all the five sub-scales to determine the suitable statistical analysis method. Based on the Kolmogorov-Smirnov and Shapiro-Wilk tests, all the sub-scales were normally distributed except for the post-instrumentality and post-learning situation scale, where $p < .05$ with a positive skew. Therefore, non-parametric tests were conducted on both

pre- and post-test of the instrumentality and learning situation sub-scales to compare the difference between the two groups over time. Similarly to the speaking data, the Mann-Whitney U test was used on the pre- and post-test, and 2-tailed statistics were reported in this analysis. This was followed by a Wilcoxon Signed-Rank test which was used to compare between pre and post-test for both groups separately. To recap, the effect size (*r*) was also calculated and interpreted in line with Cohen's *d*, where for between-group contrasts, small = .4, medium = .7 and large = 1.0; and for pre-post/within-group contrasts, small = .6, medium = 1.0 and large = 1.4. Hence, *r* values were interpreted thus: small = .25, medium = .40, and large = .60 (Plonsky & Oswald, 2014).

However, for the sub-scales which were normally distributed (attitude, integrativeness and linguistic self-confidence), outliers were identified using boxplot graphs. The results showed that all the scales were affected with outliers with a total of seven except for post-integrativeness sub-scale. These outliers provided consistently either low or high scores in both pre and post sub-scales (attitude, integrativeness and linguistic self-confidence) but were not considered extreme. To solve this, repeated measures ANOVA tests with one sample variable (group: EIST, ST) and one related sample variable (time: pre-test, post-test) were conducted on both data, with outliers and without outliers. The results were compared and the tests showed that both produced comparable outcomes. Therefore, the results from the tests including outliers are reported in this analysis.

Attitude

For the attitude scale, the result shown in Figure 5.1 revealed that there was no effect of time $F(1,57) = .53$, $p = .47$, $\eta_p^2 = .009$, or group $F(1,57) = .79$, $p = .38$, $\eta_p^2 = .02$. Likewise, there was no significant time*group interaction $F(1,57) = .19$, $p = .67$, $\eta_p^2 = .003$.

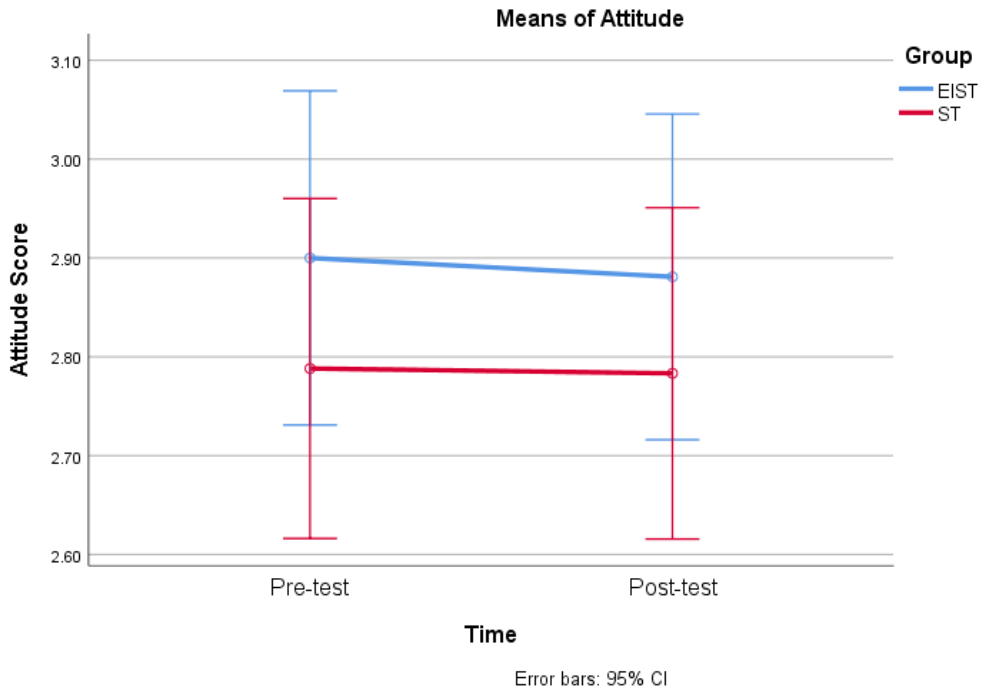

Figure 5. 1: Mean of the attitude scale over time

Integrativeness

For the integrativeness scale as shown in Figure 5.2, there was no significant effect of time $F(1,57) = .07$, $p = .79$, $\eta_p^2 = .001$, nor significant time*group interaction $F(1,57) = .56$, $p = .46$, $\eta_p^2 = .01$. However, there was a significant effect of group $F(1,57) = 6.99$, $p = .011$, $\eta_p^2 = .11$.

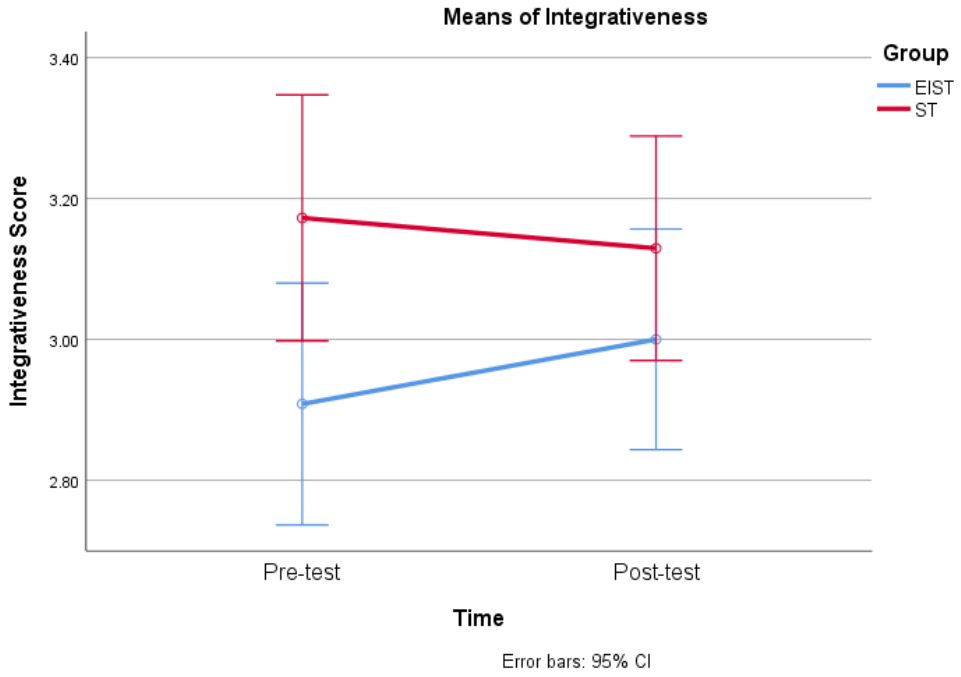

Figure 5. 2: Mean of the integrativeness scale over time

Instrumentality

Between groups: Pre- and post-test

For the instrumentality scale, the Mann-Whitney U test suggested that there was no significant difference between the EIST group (*Mdn* = 3.07) and the ST group (*Mdn* = 3.29) at the pre-test, $U = 591.50$, $z = 2.40$, $p = .17$, $r = .31$ with a small effect size. In addition, the test indicated that the scores of the EIST group (*Mdn* = 3.21) and the ST group (*Mdn* = 3.29) did not differ at post-test either, $U = 468.50$, $z = .52$, $p = .61$, $r = .07$.

Over time: Pre- to post-test

Meanwhile, a Wilcoxon Signed Rank test showed that for the EIST group instrumentality did not significantly increase between pre-test (*Mdn* = 3.07) and post-test (*Mdn* = 3.21), $z = 1.62$, $p = .11$, $r = .21$. Likewise, the instrumentality score for the ST group at pre-test (*Mdn* = 3.29) and post-test (*Mdn* = 3.29) was not significantly different, $z = .52$, $p = .61$, $r = .07$.

Linguistic Self-Confidence

For the linguistic self-confidence scale, there was a significant effect of time $F(1,57) = 6.85$, $p = .011$, $\eta_p^2 = .12$. Figure 5.3 shows that both groups made progress between pre and post-test, although the slope for the EIST group is steeper. However there was no significant effect of group $F(1,57) = .03$, $p = .86$, $\eta_p^2 = .001$ nor a significant time*group interaction $F(1,57) = 1.44$, $p = .24$, $\eta_p^2 = .03$.

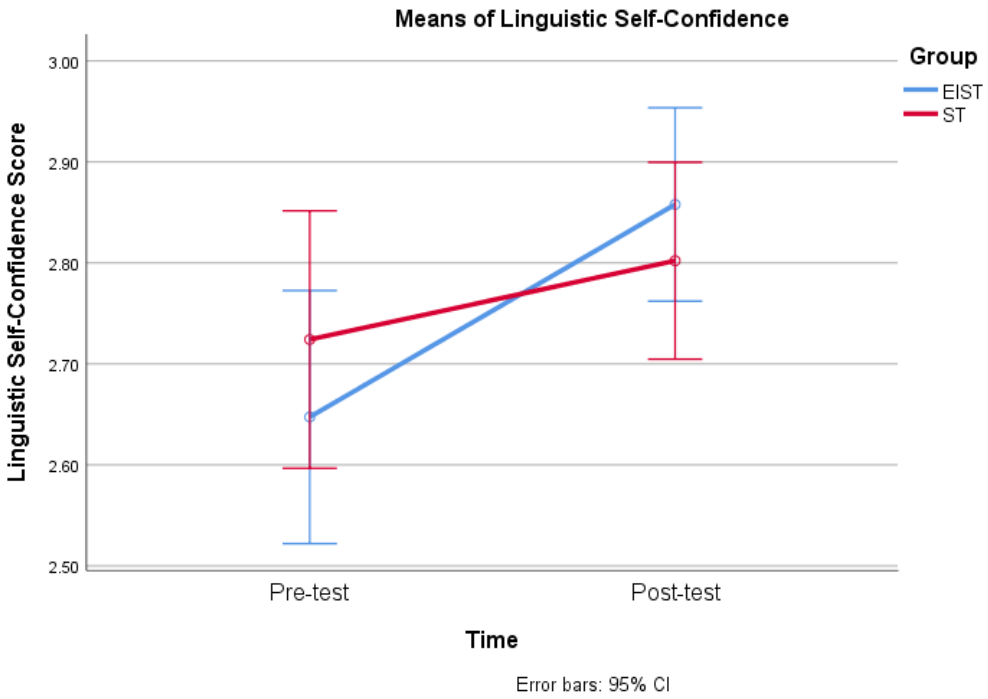

Figure 5. 3: Mean of the linguistic self-confidence scale over time

Learning Situation

Between groups: Pre- and post-test

For the learning situation scale, the Mann-Whitney U test suggested that there was no significant difference between the EIST group (*Mdn* = 2.81) and the ST group (*Mdn* = 2.75) at the pre-test, $U = 454.50$, $z = .30$, $p = .77$, $r = .04$. However, the test indicated that the score for learning situation of the EIST group (*Mdn* = 3.00) and the STgroup (*Mdn* = 2.88) differed at post-test, $U = 258.00$, $z = 2.72$, $p = .006$, $r = .35$ with a small effect size.

Over time: Pre- to post-test

Meanwhile, a Wilcoxon Signed Rank test showed that the score for learning situation for the EIST group significantly increased between pre-test (*Mdn* = 2.81) and post-test (*Mdn* = 2.75), $z = 2.83$, $p = .005$, $r = .37$ with a small effect size. However, learning situation scores for the ST group at pre-test (*Mdn* = 2.75) and post-test (*Mdn* = 2.88) were not significantly different, $z = .80$, $p = .42$, $r = .10$.

5.3 Qualitative Analysis of Motivation

Semi-structured interview was utilised to elicit how the two different forms of language teaching namely TBLT and TSLT affected students' L2 motivation towards learning the English language and speaking English. Before analysing the interview data, the researcher transcribed the interview verbatimly. In addition, the interview data were also translated from Malay to English. A copy of the transcription and the translation were sent to the students for cross check whether they agree or disagree with the transcription, or wish to add or omit any information given. This procedure is known as member checking which Creswell (2013a) described as "writing with detailed and thick description, and taking the entire written narrative back to participants" (p. 253). Member checking was considered cost-effective (Creswell, 2013a) and important because it is a form of quality control in order to improve the accuracy, credibility and validity of the interview data (Harper & Cole, 2012). Students agreed with their interview transcriptions and no changes were made.

For the analysis of the present study, thematic analysis was conducted manually. Thematic analysis is a method used to identify patterns or themes that emerged within the qualitative data (Bryman, 2012). As for the present study, the interview data were colour coded (e.g. blue to show positive view, whereas red to show negative view) and codes were highlighted in yellow as shown in Figure 5.4 below.

Main Theme	Language Level - Dörnyei's (1994)					
Sub-Theme	Attitude towards L2 Speaking					
Interview Question	EIST_07	EIST_26	EIST_22	ST_12	ST_13	ST_11
Would you avoid speaking when someone speaks or asks questions in English? Why?	Yes, most of the time. I usually bow my head down… I feel afraid to respond in English… I feel afraid people won't understand what I'm trying to say…	In school, yes! For example, I would just go to the toilet… walk around the corridor just to avoid my teacher… sometimes… when tomorrow we have to submit our homework or we have oral exam… I would just skip the class and not come to school because I don't think I can do it… I avoid it because I know I can't do it and I just feel uninterested.	No… I will speak up and express myself… but when someone asked me… of course I will answer.	Usually the teacher will ask about grammar to us… so I don't really understand… because the English grammar is confusing… so when ask to speak English… I feel like running away from class… sometimes I just skip the lesson and come back when the English lesson ended… if I can't skip class, I usually keep quiet… even if I have to speak it would be difficult for me to utter English words because most of the time Malay would come out instead… spontaneously.	When the teacher gives us exercise or ask us to answer it… mostly I just keep quiet… do nothing… because I don't feel like doing it… maybe I am not interested… I would hesitate… whether to answer it or not… but most of the time I would just keep quiet.	Before this I would avoid… usually… I would just reply I don't know… so it is easy I don't really have to say much… when I say I don't know… they would not ask further questions.

Figure 5. 4: Example of coding into theme

In Figure 5.4, several codes related to avoiding speaking and what were the students' reactions (pre-test interview) were labelled as the main theme which is language level and then organised into a sub-theme called 'attitude towards L2 speaking'. The emerging themes were coded to reflect Dörnyei's (1994) motivational framework which were allocated to one of the three levels in the framework presented below:

- **Language Level**: attitudes, integrative and instrumental motives
- **Learner Level**: anxiety and self-efficacy
- **Learning Situation**: teacher's role, willingness to communicate with other ethnics

For this study, six students were selected to be interviewed, taking three students from each group. These students were allocated a number so that their identities remained anonymous. Table 5.3 below lays out the profile of individual interviewees.

Table 5. 3: *Profile of the individual interviewees*

Group	Student	Age	Ethnicity
EIST	EIST_07	20	Malay
	EIST_26	23	Indian
	EIST_29	22	Malay
ST	ST_12	21	Malay
	ST_13	20	Malay
	ST_11	20	Malay

Sample of the students' interview transcriptions will be quoted and presented in the following section to give a comprehensive view of their responses.

5.3.1 Language Level

Motivational Orientations

At the pre-interview, motivational orientations seemed to be evenly divided between instrumental and integrative. In terms of instrumental orientation, the dominant view was that English was only worth studying to pass an examination and to get a better job, with little use for communicating with others because the Malay language was considered sufficient to be spoken. For example, some students (*N*=3) mentioned that they only interacted with non-Malay people because they avoided speaking English out of concern of being seen to show off. Most students (*N*=5) said they had no interest in attending extra English classes and two said they would only do so if they were forced to by their teacher or part of the school requirement. One student claimed that she found it difficult to connect with people when English was used. Overall, interviewees at this stage showed little sense of English being important to them.

> *"No, I don't think so because most of the time I use Malay language… I don't feel any connection when I were to speak in English."*
>
> (pre-interview: EIST_07)

> *"No, if I speak English they will think that I'm trying to show off… like… Why are you speaking in English… why don't you just speak Malay!... So I had to speak to speak Malay instead"*
>
> (pre-interview: EIST_26)

However, in the post-interview, the data showed otherwise. Students' views towards learning English language and speaking English shifted positively with the influence of integrative motives. The majority of the students (N=4) mentioned that the English language was seen to be useful to communicate with other students especially with foreign and non-Malay students, thus it created friendship and inspired them to travel outside of Malaysia and learn about other cultures. The influence of integrative motives from students' relationship with their foreign friends made them aware how important the English language is for education and also for future employment. Thus, students mentioned that they were more interested to attend English courses in the future to improve their language proficiency.

> *"Yes because in my course, I met a few foreign students… we are close… I showed them Malaysian food and they liked it… so I wanted to taste their food too… so with them I had to speak English…."*
>
> (post-interview: EIST_07)

> *"Yes because in the university… we meet different people outside of Malaysia… so I am eager to get to know them as I have never travelled outside of Malaysia"*
>
> (post-interview: ST_12)

Attitude towards L2 Speaking

Students showed negative attitudes towards speaking before the intervention was implemented. Majority of the students claimed that they did not pay attention when someone, especially their teacher, was speaking English because the level of vocabulary used in the conversation or lesson was too high for them, thus they started to lose interest and focus towards the language. EIST_26 expressed her difficulties in learning English vocabulary as she had to struggle with the Malay and Tamil language. Since English was her third language, she mentioned that English vocabulary items are mostly not phonetically uttered as compared to her L1 and L2, thus she had difficulties in remembering English vocabulary. When asked whether they were willing to ask questions and express their ideas in English, students preferred to remain reticent and would only converse with their close friends using the Malay language. Furthermore,

majority of the students claimed that most of the time, they would avoid English classes especially when speaking activities were involved:

> *'...I usually bow my head down.'* -
> (pre-interview: EIST_07)

> *"...I would just go to the toilet… walk around the corridor just to avoid my teacher… sometimes… when tomorrow we have to submit our homework or we have oral exam… I would just skip the class and not come to school..."*
> (pre-interview: EIST_26)

> *"...so when ask to speak English… I feel like running away from class… sometimes I just skip the lesson and come back when the English lesson ended… if I can't skip class, I usually keep quiet."*
> (pre-interview: ST_12)

> *"When the teacher gives us exercise or ask us to answer it… mostly I just keep quiet… do nothing…"*
> (pre-interview: ST_13)

> *"Before this I would avoid… usually… I would just reply "I don't know"… so it is easy… I don't really have to say much… when I say I don't know… they would not ask further questions."*
> (pre-interview: ST_11)

However, in the post-interview, students' attitudes towards L2 speaking positively developed throughout the whole semester in line with their integrative and instrumental motivation. The data revealed that they could pay more attention after the intervention depending on several factors. Students from the EIST group explained that they could pay more attention during the English lesson because the teacher (i.e. researcher) used simple and familiar vocabulary when teaching. In addition, EIST_29 further explained that the teacher's accent positively affected his attention towards the lesson and made him more motivated to practise speaking English. Students from the ST group added that hands-on activities made them focus more during English lessons.

Meanwhile, when asked at post-test whether they would avoid speaking, or were willing to ask questions and/or express ideas, students in both groups seemed to have different opinions, as shown below. Although the majority of the students showed a positive attitude in not avoiding speaking, they are still reserved and would restrain themselves from being expressive.

> *"I think the same… I won't respond most of the time… I will try avoid… because I'm afraid… maybe sometimes I guess… but because I just feel lazy to think."*
>
> (post-interview: EIST_07)

> *"I don't think I will avoid because I am aware how important English is… I will still try my best to answer it even though I know I am struggling… if the teacher is cool… I could just speak… but if the teacher prefers me to use English… I would force myself (to speak)"*
>
> (post-interview: EIST_26)

> *"I feel like trying… it depends on your group members… if they are outspoken… I will speak and join them… but to volunteer and talk about my ideas… a bit scared and shy."*
>
> (post-interview: ST_12)

> *"…I don't avoid it…because among my Chinese friends, I will speak English… usually it is broken English… I think the same… I would prefer to ask my friends to do it for me."*
>
> (post-interview: ST_11)

5.3.2 Learner Level

Anxiety and Self-Efficacy

The responses from the pre- and post-interviews indicated that there was a link between anxiety and self-efficacy. If the students had anxiety traits such as feeling scared, nervous, and worried or having a sense of embarrassment, they found it more difficult to speak and doubted their own abilities to make progress. In the pre-interview, all students agreed that the feeling of being anxious was caused by the pressure of people judging them when they spoke and the mistakes that they made which later affected their confidence level and their ability to speak. EIST_26 was concerned about talking to someone who had a good command of English which made her

feel intimidated. On the other hand, ST_13 admitted that her English was not good and that she made a lot of mistakes that caused her to feel nervous, embarrassed and not confident to speak. Meanwhile, ST_11 had to struggle, forcing herself to adapt to her surroundings to make herself feel worthy even though she had a mixed feeling of anxiety and being uncomfortable.

> *"If I don't know the person, I will feel embarrassed... I don't have the confidence… when I don't understand… I just say that I understand because I feel uncomfortable asking again and again….It makes me feel stupid to ask more… I do feel scared but sometimes I would force myself to say something to fit in with them."*
>
> (pre-interview: ST_11)

In consequence, students tend to use the Malay language instead of English in English lessons and conversation, but each had different views when it came to using Malay language in English classroom. EIST_07 mentioned that grammar rules in Malay are much easier because the language does not have tenses and plurality like English does.

However, dramatic changes were found in the post-interview. Although anxiety traits could not fully be eliminated in a short period of time, students in both groups felt more confident and believed they could speak and perform better:

> *"I feel less embarrassed… but if someone speaks to me in English, I will speak too. I was afraid but I do not care much anymore as long I can speak and just say what I think… I feel a little bit confident than before."*
>
> (post-interview: EIST_07)
>
> *"I am less worried… but I feel a little bit shy… but I do try to speak in English."*
>
> (post-interview: EIST_26)
>
> *"For now… I feel less afraid than before… I think when you practise and get used to it… it becomes normal… whether you do it wrongly… just do it that matters… when you mingle around with the people who are learning English… you feel motivated to speak too."*
>
> (post-interview: EIST_29)
>
> *"I feel less scared than before… when you learn more vocabulary, you feel more confident… and other students in the class have the*

same problem that I have… so we learn to improve together… so I feel less embarrassed and less scared."

(post-interview: ST_12)

"I think now I feel less embarrassed… less scared… my confidence is a bit higher than before… so I feel that I can speak a lot more."

(post-interview: ST_13)

'If people I know, I am not shy or embarrassed anymore… I can simply speak… even though I know my English is broken."

(post-interview: ST_11)

5.3.3 Learning Situation

Teacher's Role, Teaching Approach and Feedback

The results indicate that the vast majority of students believed that the teacher plays an important role in motivating them to learn and speak English. During the pre-interview, students had a negative perception towards their previous teacher which caused them not to be participative in English lessons.

"Teacher was so strict… and talked a lot which caused the students felt afraid to talk."

(pre-interview: EIST_07)

"The teacher was more talkative than the students… so most of the time we would just listen to the teacher teaching… there was a time when students talked but English was not emphasized… most of the time she will focus on the smarter students because they get things done easily… and the weak would just listen."

(pre-interview: EIST_26)

"In school, usually, the teacher will speak a lot because it eases her job… she wants to finish up the syllabus… so the students were asked to do what has been decided by the teacher."

(pre-interview: EIST_29)

According to the students, the teacher's role should be limited and students should be given the opportunity to conduct their own activity. Moreover, speaking was not emphasized in school and it was not considered as part of the lesson. English lessons were primarily focused on grammar, reading and writing as these skills were tested in national set examinations. All

students agreed that activities in the classroom were basically reading aloud, grammar drilling exercises, essay writing, and copying what is being said by the teacher or written on the whiteboard. Speaking skills, on the other hand, were only practised frequently when they entered matriculation or sixth form because it is one of the compulsory components of the MUET examination which students need to at least pass or obtain Band 3. Furthermore, ST_11 stated that in English classes, students only used English when reading and writing activities were involved but the Malay language was used as a medium of interaction among peers. Therefore, feedback by the teacher on students' speaking performance was not given which caused the students to feel that feedback was not important.

> *"… not important because like I said speaking was not emphasized."*– (EIST_07)
>
> (pre-interview: EIST_07)
>
> *"No feedback was given… because it wasn't part of the component in the exam."*
>
> (pre-interview: EIST_26)
>
> *"I speak Malay so the teacher doesn't give me feedback… because the teacher herself did not encourage us to speak… she is okay when we speak Malay in her English class."*
>
> (pre-interview: ST_11)

However, in the post-interview, these perceptions changed. Since both groups had different instruction (i.e. TSLT and TBLT), their perceptions of their teacher also differed. For the EIST group, students were more motivated to speak up and participate in speaking tasks when encouragement and guidance were given in the right amount.

> *"The teacher is not strict… the teacher taught the classroom and explained what we had to do… Students were given the chance to talk when conducting the speaking activities… they can just say whatever they like… especially when presenting in front of the class."*
>
> (post-interview: EIST_07)
>
> *"The teacher would watch us and go around the classroom… so if I don't speak I feel like I am being left out of my friends… so I start to*

> *speak on my own to make myself fit in with my friends… the teacher is aware of the student's weaknesses… so he would give extra exercises based on what we are weak in… which really helped me a lot."*
>
> (post-interview: EIST_26)

> *"I don't feel like the teacher is controlling and students were given the opportunity to express their ideas at the end of the class which I find it interesting because… I wasn't given such an opportunity."*
>
> (post-interview: EIST_29)

Meanwhile, for the ST group, even though they liked the idea of being autonomous, they still expected the teacher to give them more input on how the speaking tasks should be conducted, as they were afraid of making mistakes.

> *"The teacher did not explain that much but he asked us to present what we understand based on the video we saw… I did not know what to do… so I had to ask my friends"*
>
> (post-interview: ST_12)

> *"So the teacher only presented us the activities and we have to solve it… once we have done the task… we have to discuss it amongst ourselves… present it in front of the class"*
>
> (post-interview: ST_13)

> *"I wasn't sure what we had to do… it felt weird at first… when we had a discussion among ourselves… I was looking at the teacher hoping he could help us out and give some answers… because I want to present… I don't want to do anything wrong."*
>
> (post-interview: ST_11)

Although experiencing different types of instruction, both groups received the same feedback on their speaking performance. The feedback received seemed to bring a positive impact on the students' motivation to speak as they were beginning to realize how important speaking is, and they finally accept their mistakes and eager to make improvement.

> *"I think feedback is really important… especially after we present… because it really helped in improving my English especially using the correct grammar rules… because I am not aware of it, so I simply speak without knowing whether it is right or wrong."*
>
> (post-interview: EIST_07)

"I felt it was okay because everyone in the classroom made the same mistakes… I wasn't just the one who was wrong, but everyone was… so we were treated equally."

(post-interview: EIST_26)

Willingness to communicate with other ethnic groups and/or different background

A new situational theme that emerged in the post-interview was students' willingness to communicate with those from other ethnic backgrounds. The students believed that conversing in English with students of different ethnicity from themselves would motivate them to speak in English. The majority of the students agreed that they would not speak English to a person with the same ethnic background because it was considered awkward and to avoid being teased by their friends. Students claimed that they tended to befriend other students from different backgrounds in order to complete the speaking task because, among the same ethnic group or background, they would converse using their own mother tongue.

"If I am with Malay friends it would be difficult for me to practice English because they are reluctant to speak English… as if I speak English just to show off… their perception towards me made me not speak English."

(post-interview: EIST_29)

"When my group are all Malays… we would definitely speak in Malay and would not speak English… but I have a lot of Chinese and Indian friends, so I have to talk more in English with them."

(post-interview: ST_12)

"Chinese and Indian are mostly fluent in speaking… so when speaking it doesn't matter if you say it right or wrong… but among the Malay is different… when I meet with Malay people I would just speak Malay but when I start speaking English… they are like 'why are you speaking English, trying to show off, Malay only!"

(post-interview: ST_13)

"I have a few Chinese friends… so when I am with them I am more confident and I speak English with them… but when I meet my Malay friends automatically I speak Malay… and the task is not done properly."

(post-interview: ST_11)

The results found that students' willingness to communicate with other ethnic groups or someone from a different background emerged in the post-interview with the influence of integrative motives.

5.4 Summary

This chapter discusses both quantitative and qualitative results. The quantitative data which were analysed using SPSS 24 were presented in descriptive (Section 5.2.1) and inferential statistics (Section 5.2.2). The statistical results showed that overall, students' motivation towards learning and speaking English showed no significant difference over time for attitude, integreativeness and instrumental aspects. However, statistical results revealed that TSLT had significant improvement on the EIST group's linguistic self-confidence and learning situation motives, whereas TBLT had positive impact on the ST group's linguistic self-confidence. The qualitative data was thoroughly explored through interview and thematic analysis was used. The results showed a different picture from what was obtained from the questionnaire data. The interview analysis showed that students' motivation towards learning and speaking English gradually changed overtime. Students showed positive attitude towards learning and speaking English. In addition, students had a clear notion on the importance of English language and how it is useful to communicate with others. As for linguistic self-confidence, both groups showed high self-efficacy and managed to overcome their speaking anxiety gradually. In learning situation motive, the EIST group were more motivated as they received explicit instruction from the researcher, whereas the ST group were less motivated as they were expecting more guidance from the researcher. Therefore, the results indicated that students' level of motivation was highly influenced by their teacher (i.e. the researcher), the approach used in the classroom, and also their peers who were from different ethnic groups or cultural

background. The discussion of the results from Chapter Four, the current chapter and together with the previous literature can be found in the following Chapter Six.

CHAPTER SIX: DISCUSSION

6.1 Introduction

This chapter discusses results presented in Chapter Four and Chapter Five in relation to the existing literature on second language speaking and motivation. The discussion will be organised according to the two main research questions, which to restate are:

1. What are the effects of two different forms of language teaching (TBLT and TSLT) on:
    a. the development of students' L2 speaking performance over time?
    b. students' motivation towards learning and speaking English over time?

6.2 Overall Key Findings

The quantitative and qualitative analyses of students' speaking performance (see section 4.4 and 4.5) and motivation (see section 5.2 and 5.3) highlight several key findings. To reiterate, the EIST (explicit instruction + speaking task) group was taught using TSLT and the ST (speaking task only) group received TBLT. The statistical analysis of the speaking and motivation pre-test revealed that both groups – EIST and ST group performed at an approximately equal level prior to the intervention across all measures for speaking (lexical complexity [*G*], lexical productivity [*G*], syntactic complexity, accuracy and fluency) and motivation (attitude, integrativeness, instrumentality, linguistic self-confidence and learning situation).

However, this was not the case at the post-test. The EIST group significantly outperformed the ST group in terms of lexical complexity (*LCD* and *LCG*) and lexical productivity (*LPD* and *LPG*). By contrast, the ST group significantly outperformed the EIST group for both

breakdown fluency (*FNP*) and repair fluency (*FNR*). Nevertheless, both groups performed at an equivalent level for all measures of syntactic complexity where there were no significant differences except for the ST group which had a significant decrease for the measure of SMC (i.e. mean length of clauses). In addition, there were no significant differences in both measures of accuracy (*APE* and *ANE*). As for motivation, the EIST group significantly outperformed the ST group for the learning situation motives. In addition, both groups performed in the linguistic self-confidence motives. However, neither group made any significant change for attitude, integrativeness and instrumentality. The overall statistical outcomes for both groups are compared and summarised in Table 6.1 below.

Table 6. 1*: Comparison of results between the EIST and ST group*

	Dimensions / Measures		**Between Time**		**Between Groups**	
			EIST Group	**ST Group**	**Pre-Test**	**Post-Test**
Speaking	Lexical Complexity	*LCD*	Pre < Post***	NS	EIST < ST***	EIST > ST*
		LCG	Pre < Post***	NS	NS	NS
	Lexical Productivity	*LPD*	Pre < Post***	NS	EIST < ST**	NS
		LPG	Pre < Post***	NS	NS	NS
	Syntactic Complexity	*SMA*	NS	NS	NS	NS
		SRS	NS	NS	NS	NS
		SMC	NS	Pre > Post*	NS	NS
	Accuracy	*APE*	NS	NS	NS	NS
		ANE	NS	NS	NS	NS
	Fluency	*FNP*	Pre < Post**	Pre < Post***	NS	EIST < ST*
		FNR	NS	Pre < Post**	NS	NS
Motivation	Attitude		NS	NS	NS	NS
	Integrativeness		NS	NS	EIST < ST**	EIST < ST**
	Instrumentality		NS	NS	NS	NS
	Linguistic Self-Confidence		Pre < Post*	Pre < Post*	NS	NS
	Learning Situation		Pre < Post**	NS	NS	EIST > ST**

Note: NS = *Not statistically significant; *p < .05; ** p < .01; *** p < .001*; *LCD=Lexical diversity (D); LCG=Lexical diversity (G); LPD=Lexical density (D); LPG=Lexical density (G); SMA=Mean length of AS-unit; SRS=Ratio of subordination; SMC=Mean length of clauses; APE=Percentage of error-free clauses; ANE=Number of errors per 100 words; FNP=Number of pauses; FNR= Number of repairs.*

Within the qualitative analysis, the results showed that the language produced by the ST group was more complex, sophisticated and expressive with the presence of abstract words, stative verbs and discourse markers although less production of tokens and types was noticed. In addition, errors found in students' speech data were categorised into syntactical error (deletion of grammatical items such as auxiliary verbs, determiners, prepositions etc.); morphology (misused of subject-verb agreement and singular/plural form, determiners and word form); and lexical choice (poor choice of vocabulary). As for fluency, students were found to use breakdown fluency (filled and unfilled pauses) to assist them in thinking when elaborating ideas during speaking; and repair fluency (repetition, reformulation and replacement) as monitoring or checking whether what they are saying is appropriate and accurate.

As for motivation, students in both groups showed gradual changes in all aspects of motivation during the interview despite there being no statistical difference in attitude, integrativeness and instrumentality over time. However, the EIST group showed they were more motivated in learning English and speaking English due to the fact that the intervention had helped them to integrate with other ethnics. Meanwhile, the ST group were motivated but expected some guidance or input from the teacher in order for them to be more motivated and make fewer mistakes.

6.3 RQ1a: What are the effects of two different forms of language teaching (TBLT and TSLT) on the development of students' L2 speaking performance over time?

This section discusses the results from speaking data that provided a deep understanding of students' L2 speaking performance. The overall results presented in Table 6.1 showed TSLT (i.e. the EIST group) and TBLT (i.e. the ST group) had impacted both groups differently in all measures of speaking except for syntactic complexity (*SMA* and *SRS*) and accuracy (*APE* and *ANE*). In other words, the two forms of language teaching seem to have been effective in

promoting L2 speaking performance in their own respective ways. In the following sections, the key findings will be further discussed with respect to the effects of TSLT and TBLT on each aspect of speaking, and the interaction between the results and the measures of analysis.

6.3.1 Effects on Lexical Complexity and Lexical Productivity

There were significant quantitative and qualitative differences between each group at the post-test. For statistical analysis, the measures of *D* (*LCD* and *LPD*) and *Guiraud's Index*, *G* (*LCG* and *LPG*) were employed in this study to operationalise lexical complexity and productivity, respectively. Both measures employed provide the same statistical results, which indicates that these measures were reliable to access students' lexical complexity and productivity.

The statistical results showed that EIST group produced more diverse language with a large effect size specifically for the measure *D,* where there was no significant difference for the ST group over time. This suggests that the TSLT approach experienced by the EIST group could have assisted students in using more types of vocabulary to which they were explicitly exposed throughout the intervention period, hence greater production of word types and tokens for that group. The ST group, on the other hand, produced a lower ratio of word types and tokens. The word types and tokens were analysed qualitatively to generate an in-depth understanding of students' lexical knowledge as reflected in their usage of vocabulary. The results revealed that, although the language produced was more diverse within the EIST group, the vocabularies used were weighted with function words and less complex lexical items (e.g. concrete words) predominated, compared to the ST group whose vocabulary consisted of more complex, sophisticated and abstract words, suggesting few possibilities. There are a few possibilities that could explain the findings of this study.

First, providing form processing before meaning processing might have impacted L2 acquisition, both positively and negatively. The results of the present study are similar to a

study by de la Fuente (2006). In the previous study, a task-based lesson with an explicit FonF component was found to be more effective in promoting the acquisition of word morphological aspect than a task-based lesson that did not incorporate this component. It is possible that exposing students to grammar rules and vocabulary exemplars would have led to gains in explicit knowledge. Thus, it resulted in long-term acquisition and deeper processing of L2 vocabulary by helping students to establish more productive form-meaning connections for output production of the target words (de la Fuente, 2006; Van de Guchte et al., 2017). This explains why students in the EIST group were able to retrieve words that were taught explicitly to them and also use their own vocabulary knowledge which added to the diversity of the lexical outcome. However, the consequence was that the EIST group relied more heavily on the vocabulary practised during the intervention, thus impeded their creativity in their language outcomes. However, Van de Guchte et al. (2017) suggested that providing learners with models or exemplars may promote imitation instead of using their own creativity. Therefore, the language produced by the EIST group was repleted with concrete lexical items and function words, in spite of the diversity indicated in the quantitative type/token analysis.

Second, the inability to retrieve target words in the post-test. To reiterate, these target words were based on topics taught during the intervention (e.g., health, education, leisure and lifestyle). Since the ST group did not receive any explicit instruction nor direct guidance from the researcher, the target words were used infrequently. Hence the TBLT experienced by that group promoted shallow processing of the L2 vocabulary. This absence of guidance may constitute yet another indication that using explicit instruction (i.e. TSLT) stimulates form processing and further acquisition of words (de la Fuente, 2006). Thus, this explains why the values of *D* and *G* became lower in the speaking test for the ST group. Moreover, it appeared that the decline of word types and tokens in the ST group was caused by the presence of constantly repeated usage of discourse markers (e.g., *okay*, *right*), formulaic phrases (e.g.,

I+think, in+my+opinion, last+but+not+least) and repetitive patterns of lexis which was considered as a strategy for students to enable them to justify, explain and elaborate their ideas. Ellis et al. (2019) suggested that providing explicit instruction may cause students to treat speaking tasks as language exercises rather than communicative tasks. Thus, the speaking discourse of the EIST group was more focused on form and less on meaning. Meanwhile, the ST group was involved in more meaningful and communicative tasks that geared their speaking outcomes to become more natural and expressive. Although these students' language production was less diverse, the qualitative analysis revealed that the lexical items used were more complex and sophisticated with a higher amount of abstract words. This result is supported by Ellis et al. (2019), who discovered that the task-only group in their study produced more complex language than the explicit instruction group. A possible reason for the similarity for both the past and the present study is that these students had to use their own linguistic and non-linguistic resources without any explicit instruction prior to the speaking tasks and post-test. Hence, they had to be more independent and be creative in their own linguistic way when producing the target language.

Third, the effect of lemmatisation and the treatment given on the speaking data (see Section 4.3) might have influenced word types and tokens for both groups. This was because English is a highly inflected language and thus the lemmatisation process, which omits inflections from the analysis, is considered necessary to measure lexical knowledge rather than grammatical knowledge (Treffers-Daller et al., 2016). Therefore, the lemmatisation process was conducted on the speaking outputs of both groups to omit inflected words. The lemmatisation and the treatment overall had affected both groups' lexical outcome but not so much for the EIST group as they produced a variety of words types and tokens. In addition, the EIST group produced fewer formulaic phrases. However, for the ST group, the low production of words types and tokens were caused by the high production of formulaic phrases (e.g. *on+the+other+hand,*

last+but+not+least), in which they were treated as one type/token. This is because formulaic phrases cannot stand alone and must be treated as a single unit. Moreover, the speech discourse was truncated to 200 words for both groups which could also possibly affected the quantity of the types and tokens more, especially for the ST group. This truncation process will be explained further in the following section.

6.3.2 Effects on Syntactic Complexity

Three measures were utilised in this study to measure syntactic complexity, namely mean length of AS-unit (*SMA*), ratio of subordination (*SRS*) and mean length of clauses (*SMC*). The statistical results showed that neither group made any improvement across all measures from pre- to post-test; in fact, the ST group had a significant decrease for the measure of *SMC*. There are a few possible explanations for these findings.

First, as mentioned earlier, both pre- and post speaking tests for both groups were truncated to 200 words (tokens). It could be hypothesized that cutting down the words to 200 may have led the values of these measures to decrease especially the measures of *SMA* (total number of tokens / AS-units) and *SMC* (total number of tokens / total number of clauses). Therefore, when the number of tokens for all speaking tests was reduced to 200 words, all students would have approximately the same outcomes in which the average scores for AS-units and the total number of clauses were 15 and 28, respectively. However, the process of truncation was considered necessary as it was done to harmonise all the global measures of speech production. It could be assumed that students' syntactic complexity was poor due to the effect of the truncation process and not because of students' incompetence. If the speech production was not reduced to 200 words, perhaps they could have made better progress. However, the truncation process was done to prevent other variables (e.g., lexical complexity, speech length etc.) from exerting influence on the dependent variables (Treffers-Daller et al., 2016).

Second, the complexity of the tasks affects syntactic complexity. The AS-unit is used to measure speech and it consists of "an independent clause or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000, p. 365). In other words, a complex sentence is described as having two or more clauses. In the qualitative analysis (see Section 4.7.1), however, the results showed that within a single utterance, subordination was produced occasionally and mostly contained one main clause only. Therefore, students used more simple sentences which only contained one independent clause and no subordinate clause which affected the measures of SMC and SRS (total number of clauses / AS-unit). It could be assumed that the speaking tests were not complex or challenging enough to elicit students' syntactic knowledge. The findings of this study are not in line with the study conducted by Mochizuki and Ortega (2008), which found that students who engaged in guided planning produced significantly more relative clauses in their oral narrative tasks than those who engaged in unguided planning. The study suggests that greater demands on guided planning are associated with more complexity in terms of the mean number of relative clauses per T-unit. The tasks in the previous study were guided planning oral narrative tasks which involved students working in pairs. The guided planning involved a picture storytelling task with an additional handout which briefly explained relative clauses during the five-minute preparation time. Therefore, the pair-work oral tasks with guided planning may have prompted students to formulate more complex structures resulting in more syntactically complex language with the increases of relatve clauses. However, this is not the case for this present study in which every student was assessed individually and no guidance (e.g. handout) was given to them in both speaking tests, hence the speaking tests were less effective in inducing more complex syntactic structure.

6.3.3 Effects on Accuracy

Two measures of accuracy the *percentage of error-free clauses* (APE) and *number of errors per 100 words* (ANE) were employed to examine whether the implementation of TSLT and TBLT would affect speaking accuracy. The descriptive statistics showed that the EIST group outperformed the ST group in the post-test on both measures of accuracy. However, although the EIST group showed improvement, the pre-post change was marginal and indeed a 2x2 ANCOVA revealed that neither group made significant progress. Thus, neither TSLT nor TBLT approaches resulted in an increase in error-free clauses. In fact, students produced more errors per 100 words. A thorough analysis was done to compare why such an effect occurred in both groups following the intervention. Several assumptions derived from these results.

First, students' primary attention was drawn to content and meaning especially for the ST group. This could be seen in the speaking outcome of the ST group which showed that the type of language used was heavily weighted towards formulaic phrases and discourse markers (see Section 4.7.1). Therefore, it could be considered that using TBLT which only employed speaking tasks without the explicit instruction led to FonM. This is in line with the study by S. Li et al. (2016) that compared TBLT and TSLT. Their findings revealed that TBLT led to limited opportunities for learning the target language and thus discouraged deeper cognitive processing of the grammatical forms and structures to take place (S. Li et al., 2016). In the present study, it could be assumed that the speaking tasks provided insufficient experience of the target structure for the ST group which resulted in little acquisition, although opportunity for learning was given. Therefore, this could lend support to the argument that noticing is a necessary condition in order to learn language form (Schmidt, 1990, 2010). Meanwhile, the non-significant statistical results for the EIST group in the present study might be caused by a trade-off effect between achieving the target structure and using the vocabulary exemplars. It could be assumed that students were 'testing out' a wider range of vocabulary at the post-test,

and thus failed to achieve the target structure, resulting in overall accuracy to decrease. Further discussion will be given in Section 6.3.5 regarding these effects on both groups in relation to the Noticing and Limited Attention Capacity Hypothesis.

Second, the feedback was given to the students in both groups. Although the main focus of the study was mainly on the two different forms of language teaching (TSLT and TBLT), feedback was given as it is a part of the course requirement to evaluate and give students feedback on their performance. However, feedback was not done rigorously to avoid overshadowing the effects of the different forms of language teaching (TSLT and TBLT). The nature of the feedback in both groups was the same in the during-task stage when the speaking tasks involved pair work activity and group discussion. During this stage, both groups received confirmation checks without providing them with any correction. Confirmation checks were given spontaneously when the researcher was observing the class, group by group. Meanwhile, in the post-task stage, additional feedback was given. In this stage, both groups were required to do a group presentation from what had been discussed in the during-task stage. Feedback was given after all groups had presented rather than during presentation to avoid repeated interruption which could impede the flow of the presentation. A few errors were selected randomly from the students' presentation and explicitly explained to the classroom as a whole, rather than as an individual oral assessment since time was limited. S. Li et al. (2016) revealed that groups that received attention to the form (explicit instruction and recasts feedback) acquired the target language (passive verb forms) effectively and explicit instruction had a greater effect on students' language accuracy. However, the findings of the previous study do not corroborate those of the present study. The students in both groups of the present study did not make progress in accuracy.

Third, the characteristics and conditions of the task. The lack of improvement in accuracy might be arguably be because the nature of the tasks used in this study, which was different

compard to the previous study. In S. Li et al. (2016), dictogloss tasks were conducted by the teacher and students were required to listen to the narrative presented, work in pairs, rehearse with no time limit and retell to the class. However, in the present study, the time given for planning and completing the speaking tests was short. In addition, the speaking tests were conducted and assessed individually, thus students did not receive any guidance nor did they have time rehearse. Although the feedback given during the intervention did not fully benefit students' accuracy in the post-speaking test, it did somehow give students opportunities to reflect on and modify their linguistic knowledge especially for the EIST group when they made more reformulations over time (see Section 6.3.4 for discussion in relation to reformulation and students' fluency outcomes). Therefore, the explicit instruction and feedback may have prompted students in the EIST group to notice the gap between their existing knowledge and target-like knowledge, hence prompting them to reconstruct existing knowledge towards the target-like representations.

Fourth, fossilisation may have occurred. The qualitative results showed that errors concerning morphology (e.g., verb tense, subject-verb agreement and word form) constituted more than half of the total number of errors. Moreover, these errors seemed to be persistent and frequent over time, hence providing evidence of possible fossilisation in specific language areas in spite of the intervention. The findings of this study are not in harmony with the study of Van de Guchte et al. (2017), which revealed that the FonL group (i.e. focus on language, given written instructions directed to language) were able to achieve the target structure more accurately than the FonC group (i.e. focus on content, given written instructions directed to content). The contradictory findings of both previous and present study might be because of the students' age and language experience. The average age of the students in the previous study was 14 and the German language was newly introduced to them as a foreign language in secondary school (two hours per week for about 17 months). However, the average age of the students in the

present study was 21 and English had been taught to them for nearly 13 years since primary school. It is possible that students in the present study, who had learned English grammar for some time, had committed errors during English lessons in the past which were not properly corrected, leading to fossilisation, whereas students in the previous study just had started to acquire the new language.

6.3.4 Effects on Fluency

Two measures were employed in this study to operationalise fluency: number of pauses (*FNP*) and number of repairs (*FNR*). The quantitative results revealed that both groups, the EIST group and ST group, managed to improve breakdown fluency by reducing the number of pauses from pre- to post-test with small and medium effect sizes, respectively. In contrast, for repair fluency, only the ST group significantly reduced the overall number of repairs in the post-test with a small effect size (see Section 4.6.2 for fluency results). It should be noted that the characteristic of the tasks in the speaking tests (pre- and post-tests) were different from those used in the intervention. The questions in the speaking tests were designed as a monologic task to assess students' overall speaking performance individually, whereas the speaking tasks throughout the intervention period were dialogic in nature. The conditions in which both tasks were performed also differed and need to be taken into consideration. The speaking tests had a specific time limit for students to perform and planning time was given prior to the tests. In the intervention, however, time was not restricted and ample time was given for group discussion and pair work. Therefore, this allowed for an existing speech gap between the interlocutors to comprehend information, plan their speech systematically and negotiate meaning with their peers using both linguistic and non-linguistic resources effectively.

A thorough analysis was carefully conducted in order to find pausing patterns. The results were that both groups produced more filled pauses than unfilled pauses in both pre- and post-tests.

The filled pauses (e.g., *uh, um*) in both groups were highly used in mid-clauses of the speech and frequently occurred after a repair was made. Meanwhile, unfilled pauses or silent pauses were found to be more prominent in both mid-clauses and at the beginning or end of the clauses. For repair, both groups seem to have approximately the same amount of repetition (repeated words and/or phrases) and replacement (word and/or phrase replacing) from pre- to post-test. However, the major difference was in reformulation (correcting words and/ or phrases) between both groups. The ST group drastically reduced the amount of reformulation from pre- to post-test compared with the EIST group which had a small decrease. Therefore, these findings indicated that speaking tasks without the influence of explicit instruction had more positive effects on students' fluency. There are possible explanations for these effects.

First, the occurring pauses might be caused by linguistic or cognitive processes. A study by Tavakoli (2011) investigated the differences in pausing patterns between L2 learners and native speakers in four oral narrative tasks (picture-story narration). The results of that previous study suggested that L2 speakers produced more unfilled pauses in the mid-clauses and paused less at the end of the clauses across all four tasks. Similar results were found in the present study where the unfilled pauses occurred in mid-clause. As for repairs, Tavakoli (2011) suggested that L2 speakers paused before they repeated a vocabulary item (repetition), replaced a word or a phrase/expression with another (replacement) and/or prior to correcting (reformulation). In the present study, these pauses were inconsistent, depending on the students in which pauses could occur both before and after repairs. This indicates that students in both studies used pauses when they were thinking to change or reshape the structure of an utterance (repetition and replacement) and when monitoring their own performance or checking whether they were using accurate and appropriate language (reformulation) (Tavakoli, 2011). Furthermore, in the present study, the speaking data indicated that some pauses occurred during online planning. From this it could be assumed that pauses helped students when they were formulating their

thoughts, organising ideas and figuring out how to convey the message into words. Another interesting fact about the current findings which is similar to the previous study is that students paused outside the formulaic sequences and rarely in the middle of a formulaic sequence (e.g., *last+but+not+least, in+the+meantime, and+so+on*). It could be assumed that these chunks of words require less attention and no additional processing is needed (Tavakoli, 2011) and thus they were 'ready to go' and were uttered automatically by the students with ease.

Second, the level of proficiency might have affected their fluency outcome. The students in the present study were considered as intermediate second language learners of English and currently in their first year of undergraduate study. The positive effects of the intervention on fluency especially breakdown fluency in this study differ from those in Ellis et al. (2019). The previous study showed that students were having difficulties in producing fluent speech, with more frequent and longer pauses, especially in the group which received explicit instruction. This might be arguable because, in the previous study, English was considered as a foreign language where students do not use the language frequently in their daily life. Moreover, the students in Ellis et al. (2019) were secondary school students and the average age was 14 years. However, in the present study, students were in their tertiary level and English proficiency level was considered intermediate, and thus their experience perhaps diminished the negative effect of increasing demands on fluency.

Third, explicit instruction affects fluency performance. Although pre-task planning was given, the EIST group did not make any significant decrease in reformulation. This suggests that the EIST group's attention was more on accuracy in which they were keen to correct themselves in order to produce accurate language and meet the target structure. This is in harmony with Ellis et al. (2019) who argued that the explicit instruction group concentrated more in producing the grammatical target structure, which explains why students paused more regularly and for a longer period. However, the previous study did not look into repair fluency. Although it is

arguable that repair and breakdown fluency are different, such analysis could reveal that the students who received explicit instruction in both studies were struggling with their fluency.

6.3.5 Effects on Speech Production in relation to Noticing and Limited Attention Capacity Hypothesis

According to Skehan (2009), successful performance in a task-based context has often been characterised as containing "more advanced language, leading to complexity; a concern to avoid error, leading to higher accuracy and the capacity to produce speech at normal rate without interruption, resulting in greater fluency" (p. 510). If this was the case, students in this study did not meet the criteria. The results of the present study showed that both groups performed differently and were not able to perform in all aspects of speaking simultaneously. Building upon the discussion, several assumptions are reviewed below which focus on the interrelationship between complexity, accuracy and fluency within the Speech Production Model (Kormos, 2006; Levelt, 1989), and also the Limited Attention Capacity (Skehan, 2009) and Noticing Hypothesis (Schmidt, 1990).

First, processing L2 is not the same as L1. As mentioned in Chapter Two, speech production models for L1 (Levelt, 1989) and L2 (Kormos, 2006) were used in the study. Generally, both models have the same three main processes (conceptualisation, formulation and articulation) of speech production, but these processes differ in how speakers produce their speech for L1 and L2. The L1 speech production stages occur in parallel and are efficient because the mental lexicon is extensive and well-organised (Skehan, 2015). However, this is not the case for the L2. The ability to conceptualise information in L2 does not differ from L1 but formulation is more vulnerable in L2 speech production. This is because of the L2 learners' insufficient mental lexicon resources which require more effortful syntax building processes (Skehan, 2015). Moreover, serial processing occurs in L2 speech production, and thus greater demands are

made on limited attentional resources to repair the problems encountered at the formulation stage (Skehan, 2014). Therefore, it is assumed that complexity is established by the conceptualiser during the message planning process, whereas accuracy and fluency are generated by the formulator during the process of translating the conceptual message into a linguistic structure (Kormos, 2006; Levelt, 1989).

Second, the importance of awareness and noticing linguistic features. The involvement of the mental lexicon depends on how much information it receives from the conceptualiser in a certain time frame that is associated with the capacity to produce fluent and accurate language (Skehan, 2015). Based on the accuracy results, it could be assumed that students failed to retrieve the correct lemma and connect them with the correct syntactic structure(s) during the formulation process. This could have been a challenge for the EIST group as students produced more reformulation, proving that they were aware of the incorrect usage of grammatical items and were trying to correct the errors committed. However, this seemed to be less the case for the ST group, who, when completing the speaking post-test seemed to be solely focused on meaning as is reflected in the low score for reformulation. It could be assumed that the ST group may have had less awareness of grammar rules and the mistakes made. Therefore, they might see accuracy as less or not important, as the grammar rules were not stressed or taught explicitly. This also lends support to the Noticing Hypothesis which claimed that learners must attend to and notice linguistic features of the input that they are exposed to if those forms are to become intake for learning (Schmidt, 1990). Therefore, if the ST group had been given explicit instruction, it might have caused them to have an awareness of the target structure being tested and pay more attention, resulting in more noticing and persistently understand the significance of noticed language features (Schmidt, 2010).

Third, attention and working memory capacity are limited, depending on the condition of the tasks. The results showed that both groups performed better in complexity (except for syntactic

complexity) than accuracy. When the task is demanding, which in this study, the time limit of the speaking tests, students attentional resources are not enough to cater to both meaning and form at the same time. This agrees with Skehan (2009) who argued that resources are divided between accuracy and complexity when a natural tension is present. This could be supported by the Limited Attention Capacity Hypothesis (Skehan, 2009, 2015). Students were not able to focus on all aspects of speaking during the tests, as their attention and capacity were weighted more on planning and meaning in the conceptualiser. Skehan and Foster (2012) suggested that complexity and fluency are linked together quite frequently and similarly for accuracy and fluency, but it is unlikely, though not impossible by any means, for complexity and accuracy to be increased at the same. The findings of Van de Guchte et al. (2017) suggested that when the FonL group in their study focused on the grammatical structure during task performance, it may have had a negative impact for the complexity of their task performance. The opposite effect occurred in the present study in which the EIST group performed better in lexical complexity but their accuracy score decreased. Therefore, it could be assumed if one aspect is elevated, the others tend 'to be sacrificed' due to the attentional limitations to meet the characteristics and conditions of the task.

6.4 RQ1b: What are the effects of two different forms of language teaching (TBLT and TSLT) on students' motivation towards learning and speaking English over time?

This section discusses the results from the questionnaire and interview that provided valuable insight into the students' L2 motivation. The variety of motivation components in Table 6.1 shows that "L2 motivation is an eclectic, multifaceted construct" (Dörnyei, 1994, p. 279). The overall results showed that the EIST group became more motivated than the ST group in terms of the learning situation motives, which indicates that TSLT might have positive effects to a certain degree. In the following sections, the key findings will be further discussed with respect

to the effects of different forms of language teaching (TBLT and TSLT) on L2 motivation from the perspective of three levels, to correspond with three basic constituents of the L2 learning process: language level, learner level and the learning situation.

6.4.1 Language Level: L2 Motivation and the Role of English

Within the language level, the focus is on the orientations and motives related to basic learning goals and the potential usefulness of proficiency (Dörnyei, 1994). In the present study, the statistical results in both pre- and post-test showed that there was no significant effect on both groups for instrumental and integrative motives. The qualitative data, however, showed otherwise.

In the pre-interview, students in both groups had only vague notions that English language learning, specifically developing the speaking ability, is a good thing to do. In addition, they were unable to articulate clearly their main reasons behind learning and speaking English. Students had just started their tertiary level and were fresh school leavers. It could be assumed that students were not convinced about the importance of English for future purposes especially the need to speak English in a community where the Malay language is widely used. Therefore, they were perhaps still confused about the role of English and what it meant to them in the future. However, in the post-interview, all students mentioned more concrete explanations of what was the purpose of learning the target language (e.g., passing exams and career-focused) and asserted their awareness about the role played by the English language in current society, which led students to be more instrumentally motivated. A similar pattern was observed by Hong and Ganapathy (2017). They conducted a qualitative study to identify whether instrumental or integrative motivation plays a more important role in promoting ESL students' English language learning. Interviews were conducted with ESL students and it was discovered that students were more instrumentally motivated than integratively motivated. The study

suggested that students in certain contexts learn English for the sake of passing an exam and getting better jobs in the future.

As for integrativeness, the qualitative analysis revealed that students from the EIST group claimed that, besides the intervention, the presence of ethnic groups other than Malay and foreign students whether in or outside the classroom made them feel motivated. Therefore, it seemed that the EIST group was integratively motivated. The EIST group was more positively disposed towards the idea of integration with the target language community (i.e. English). In Malaysia, alongside the Malay language, English plays a major role in all contexts (e.g. education, administration, leisure etc.). In addition the language acts as a lingua franca that connects people from other different backgrounds especially with the presence of foreign students in their university course. Thus, it appears that integrative motivation to learn and speak English among the students from the EIST group was relevant in this context in order for them to assimilate themselves with other ethnic groups and/or foreign students especially in an environment that involves group interaction and group work. The qualitative findings of this study are also in line with Hong and Ganapathy (2017) who also discovered that integratively motivated students were more motivated to learn the target language as they participated in more classroom activities (e.g. doing homework, reading aloud etc.) compared to those were instrumentally motivated who did not commit much to their English language learning. Although students in the previous study were instrumentally motivated and aware of the importance of English, they did not put much time and effort into mastering it compared to the students who were integratively motivated.

6.4.2 Learner Level: Linguistic Self-Confidence

The learner level involves an intricate relationship between affects and cognitions that form fairly stable personality traits (Dörnyei, 1994). In this study, two motivational components

underlie the motivational processes at this level, need for achievement and self-confidence. The latter encompasses language anxiety and self-efficacy which are the main focus of the study. The statistical results showed that both groups had a significant improvement in linguistic self-confidence from pre- to post-test, although the change was greater for the EIST group. This suggests that the intervention had a positive effect on both groups but the explicit instruction affected the EIST group's linguistic self-confidence to a larger extent.

In the interview, students said it felt strange to speak English among peers when everyone understood Malay well. Therefore, English was not spoken nor practised in a consistent manner even in English classroom settings which led to negative opinion from the students throughout the pre-interviews. Students described the classroom as an anxiety-inducing environment in which they became reticent, nervous and in fear of making mistakes. Moreover, students believed that due to their limited English grammar and vocabulary, they doubted their own ability to speak English effectively in front of others. However, in the post-interviews, positive changes were found where students described themselves as being less nervous, more confident in speaking and believed that they could speak and perform better depending on external factors (see Section 6.4.3). As the interview questions pertaining to linguistic self-confidence were geared towards the ability to speak (i.e. speaking self-efficacy), students' responses mostly referred to their ability to speak such as speaking confidently without the fear of making mistakes, giving presentations in front of the class without feeling nervous.

The findings from the pre- and post-test interviews could relate to the other findings concerning self-efficacy. This study substantiates the findings of Mohamed Khatib and Maarof (2015) who conducted a study on two different groups (first and fifth-semester) of Malaysian polytechnic college students, looking at their self-efficacy for oral communication. The results revealed that fifth-semester students had a high level of self-efficacy in speaking due to the fact that they had enrolled in several English courses in their previous semesters and completed their

industrial training. Hence, they were exposed to authentic use of language in the 'real world' and had vicarious experiences during their industrial training (i.e. another recognised source of self-efficacy (Bandura, 1997)), which seems to have strengthened their self-efficacy beliefs in speaking. By contrast, the first-semester students' self-efficacy was distinctly low and they tended to struggle to complete speaking tasks which led to frustration. This suggests that, in the present study, students' low levels of self-efficacy in the pre-interview of the current study may be attributable to the fact that they were in their first semester, and were taking FoEL (Fundamental of English language) course which is an English remedial course, compulsory to pass for students who achieved Band 3 and below for the MUET examination. Therefore, the intervention was designed not just to meet the requirement of the course but also improve their English by allowing them to gradually build up their speaking skills in order to take part in various speaking tasks (e.g., group discussion, presentation, story-telling, role-play etc.). In the course, students in both groups of the present study were newly exposed to various types of communication skills, experiences which rarely occurred in their primary and secondary school. Bandura (1997) emphasised that mastery experiences are important for the development of self-efficacy. Therefore, students had had insufficient exposure, experience and knowledge on speaking activities, and were still in a phase of exploring their own strengths and weaknesses. However, in the post-interview, students' self-efficacy gradually increased as they claimed to be more confident in speaking and believed that they were able to speak English. This shows that throughout the intervention, students in both groups had positive experience and the teacher (i.e. researcher) might have influenced them positively, could also be a reason.

6.4.3 Learning Situation: The Influence of External Factors

The statistical results showed that the EIST group made a significant improvement in learning situation motives with a small effect size form pre- to post-test, whereas the ST group showed no significant change. Qualitative analysis was conducted to gain an in-depth explanation of the differences between both groups and what caused these effects to occur. This section concerns extrinsic motives and motivational factors which emerged as prominent themes throughout the interview, namely the teacher's role and the type of instruction employed in the classroom and willingness to communicate with other ethnic groups.

6.4.3.1 Teacher's Role and Instruction Employed

Teachers and the methods used in teaching play a significant role in creating a better environment to allow language learning and facilitating students' ability to overcome learning obstacles (Dörnyei, 1994; Dörnyei & Ushioda, 2011).

A qualitative analysis was conducted to obtain an in-depth explanation of this effect on students' motivation. In the pre-interview, the majority of students expressed negative perceptions towards their teacher and how the lesson was conducted. To summarize, students mentioned that the class was teacher-centred and speaking was not emphasized as important because the lesson was geared towards learning grammar, reading and writing, which were basically tested in most English examinations in Malaysia. However, these negative perceptions gradually changed in the post-test. The students from the EIST group felt more motivated when guidance was given sufficiently and simultaneously giving them opportunities to speak and participate in speaking tasks. Nevertheless, although the ST group liked the idea of being autonomous, they still expected the teacher (i.e. the researcher) to give them some direct input as they were afraid of making mistakes.

The findings of this study support those of a recent study by de Smedt et al. (2018), which explored the effects of explicit instruction on students' writing motivation. The study looked into the effect of explicit instruction and peer-assisted writing on students' writing motivation and self-efficacy. 206 elementary school students were grouped into four groups, namely explicit instruction + writing individually (EI+IND), explicit instruction + peer-assisted writing (EI+PA), peer-assisted writing (PA) and individual writing (IND) only. The previous study revealed that students who received explicit instruction had higher motivation after the intervention but they were dependent on the teacher (i.e. controlled motivation). This indicates that the explicit teaching of writing may have increased students' writing motivation but in a less favourable way. The instruction used probably hindered students' spontaneous writing, which later resulted in students feeling unable to write and use the writing strategies taught. In the present study, this may be the case for the EIST group as indicated by their speaking outcome where they demonstrated heavy use of target words which could indicate that the EIST group had a sense of dependency even though they were highly motivated. Meanwhile, students in the ST group in the present study were very autonomously motivated at the post-test, as were the PA students in the previous study. De Smedt et al. (2018) suggested that peer-assisted writing encouraged students' innate needs, hence creating autonomous writing motivation. The previous study proposes that allowing students to work with peers increases their motivation to write out of satisfaction, pleasure or recognition of the value of the writing tasks (de Smedt et al., 2018). However, the previous results only partially agree with the present study. Students in the ST group did experience autonomous learning especially when they had to complete the speaking tasks collaboratively during the intervention period. Consequently, it did enhance autonomous speaking motivation and complex speech. However, students still expected some input from their teacher (i.e. the researcher) because working with peers seemed not to have been enough and making errors were not considered ideal to them. This could explain one of

the reasons that there was no significant improvement in learning situation motive for the ST group.

6.4.3.2 Willingness to Communicate (WTC) with Other Ethnic Groups

A person's WTC can be influenced not only by personality-traits but also by situation based variables such as intergroup motivation; a certain behaviour towards members of a particular group (MacIntyre et al., 1998). Since Malaysia is considered a multiracial and multicultural country, a theme surfacing from both pre- and post-interview relates to students developing a sense of self-awareness especially in relation to ethnicity.

In the interviews, the Malay students in both groups provided unfavourable responses when asked about speaking English with their peers from the same ethnic background. A sense of resentment amongst the Malay students was noticeable when English was used or spoken because the language itself did not suit their linguistic identity and was not seen to fulfil their daily communicative purposes. The Malay students felt that they did not sound 'themselves' and were less expressive when speaking English and were not able to 'fit in' within their community. There are possible assumptions underpinning the students' attitudes in the present study. The findings of this study are in line with a study by Abdul Razak et al. (2018) who examined the relationship between ethnic group affiliation (EGA) and WTC. In the previous study, EGA was identified as how closely a person feels attached to their native background, culture and language (Abdul Razak et al., 2018). The previous study suggested that the ethnic groups often pressurize their own members to behave or act in a certain way in order to preserve their existence, language, culture and identity which resulted in students being unwillingly to speak other languages than their own. To some extent, this could be the reason why students in the present study would favour the Malay language to communicate among themselves even in English language classroom or other English language settings in order for them to be accepted by their 'own people' and avoid being isolated.

However, the Malay students felt more confident and motivated to speak English with other ethnic groups than their own because English speaking tasks could be completed in a timely manner through the use of the target language. This also could be caused by students' integrative motivation, especially for the EIST group. However, within the same ethnic group, speaking tasks could not be accomplished because of the use of their mother tongue amongst them. Students recognised the functions and importance of English and as a necessary tool to achieve task completion which led back to instrumental motivation. This outcome agrees with one of the variables in the outcomes of S.-J. Kang (2005), whose study was carried out to investigate how situation-based variables can influence Korean students' WTC in the students exchange programme in the United States. Through qualitative analysis, the results showed that the WTC was caused by three main variables: excitement, responsibility and security (S.-J. Kang, 2005). The previous study suggested that students felt excited to communicate when topics were related to their personal experience (e.g. family background, culture, language). Moreover, a sense of responsibility emerged when students have to clarify the wrong information and seek useful or important information. In addition, students felt insecure and unwilling to speak if the situation involved many interlocutors and speaking with non-native speakers who were fluent in English. As for the present study, students mentioned that they were more focused on the speaking tasks and were eager to finish them within the given time. Therefore, there was a sense of responsibility to participate in the speaking tasks with other ethnic groups and to complete the tasks. However, the present study did not find a sense of excitement and security in students' interviews. It could be assumed that the majority of the students came from the same background and have the same language proficiency. In addition, the difference between the two studies may spring from the fact that the study by S.-J. Kang (2005) was not an intervention, while the present study was. Therefore, it could be hypothesised that the design of the study and how it was conducted influenced students' main intention which

led to different outcomes. The students in the present study were motivated towards task completion as they participated in the speaking tasks.

6.5 Summary

This chapter discussed the results of the findings from Chapter Four and Five in line with the research questions. The structure in this chapter does not imply that the discussion can be neatly segregated and contained within a single research question or method which relates to it. The aims of the study were to investigate how two different forms of language teaching (TBLT and TSLT) affect speaking performance and L2 motivation over time. Therefore, the results arising from the implementation of two different forms of language teaching were carefully linked by the research question to show how they relate to speaking performance and L2 motivation. Each type of approach implemented in this study contributed a small amount of new information which gives insights into how types of language teaching could affect students' speaking performance and L2 motivation differently. The next chapter draws the implications of the discussion together; it reflects on the research aim of the study, explains the contribution, outlines limitations and suggests directions for future research.

CHAPTER SEVEN: CONCLUSION

7.1 Introduction

This final chapter presents the conclusion to the present study. The summary of the study and findings are presented first, followed by the limitations of the study and its contribution to knowledge. The chapter also outlines recommendations for future research.

7.2 Summary of the Study

This study aims to investigate the effects of two different approaches, namely task-based language teaching (TBLT) and task-supported language teaching (TSLT) on students' speaking performance and motivation towards learning and speaking English. The main difference between the two types of instruction is that TSLT attempts to integrate explicit instruction within tasks based on specific linguistic features in a more complex pedagogic setting (Skehan, 1996). In contrast, TBLT rejects the role for explicit instruction prior to task performance (Long, 2015). The rationale behind the present study was derived from students' low level of speaking performance, and their lack of motivation in communicating in English. This was hypothesised to be the result of their insufficient exposure to activities designed to develop speaking skills, in a context where pre-university English language education is mainly focused on reading and writing skills.

The study adopted a mixed-method, quasi-experimental, pre-post-test design. Two intact classes were assigned to two groups in this study, namely the EIST (explicit instruction + speaking tasks, i.e. TSLT) group and the ST (speaking tasks only, i.e. TBLT) group. The participants were 59 English as a second language (ESL) undergraduate students (aged 20 to 23) from a technical / engineering university in Malaysia. They were randomly assigned to the

EIST group (N = 30) and the ST group (N = 29). The intervention was conducted for two hours weekly for a duration of eight weeks.

The instruments of this study involved both quantitative and qualitative methods. Students' speaking performance was measured by speaking tests which were conducted individually. Meanwhile, students' level of motivation towards learning and speaking English was elicited through the use of questionnaires and interview. The instruments were administered at two different time points: at the pre-test, in order to gain insight into students' existing level of speaking performance and motivation before the implementation of the intervention; and at post-test, in order to measure whether the intervention had influenced students' speaking and motivation. For the data analyses, eleven measures for speaking were used to determine the effectiveness of the intervention regarding speaking complexity, accuracy and fluency. In addition, five sub-scales for motivation covering attitude, integrativeness, instrumentality, linguistic self-confidence and learning situation were utilised to measure students' overall L2 motivation at each time point.

7.3 Summary of the Findings

With regard to RQ1a (the impact on students' speaking), the data analyses showed that students in both groups had almost an equal level of speaking performance at the pre-test. However, in the post-test, the results showed that the effects of the intervention on students' speaking performance differed between the two groups. The analyses indicated that the TSLT approach employed in the EIST group had a significant effect on students' vocabulary outcomes, whereby the language produced was more diverse than the students in ST group, which resulted in higher scores for the measure of lexical complexity and lexical productivity. As for the ST group, the TBLT approach led to no statistically significant increase for either lexical measure over time but the qualitative analysis indicated that students were able to produce more

complex and sophisticated words (i.e. abstract words) compared with the EIST group which produced more concrete words. In addition, the TBLT approach employed in the ST group seemed to be effective in improving students' fluency with a decrease in breakdown fluency (e.g. pauses) and repair fluency (e.g. repetition, reformulation) compared to the EIST group. Nevertheless, neither group illustrated a statistically significant improvement across all measures of syntactic complexity and accuracy.

As for motivation towards learning and speaking English (RQ1b), the data analyses revealed that both groups had approximately an equal level of motivation at the pre-test. However, the questionnaire data indicated that the two approaches had no significant effect on students' attitudes, integrativeness or instrumentality over time. Although students did not make any significant improvement from pre- to post-test, the analyses of the post-interview confirmed that students, especially in the EIST group, showed more positive attitudes towards speaking English and had a clearer notion of what integrativeness and instrumentality meant to them than at pre-test. An awareness of the importance of learning and speaking English seemed to lead students in both groups to be more instrumentally motivated at the time of the post-interview. As for linguistic self-confidence, both teaching approaches had a positive effect especially in the case of the EIST group. In that group, students felt more confident in speaking and believed that they were able to speak in English without feeling nervous. Meanwhile, for the learning situation motives, the EIST group made a significant improvement from pre- to post-test whereas the ST group did not. The results from the interview analysis indicated that the teacher's role, especially in providing explicit instruction, and the influence of other ethnic groups were the main reasons why students in the EIST were more extrinsically motivated than the ST group.

7.4 Limitations of the Study

Limitations to the present study are acknowledged. First, the present study adopted a quasi-experimental design in which students were not randomly assigned to a specific group based on certain criteria (e.g. English language proficiency, gender, undergraduate major etc.). Two intact groups were assigned by the language centre as the researcher had no control over students' enrolment in the language course. Therefore, it was decided that the two groups would remain intact rather than allocating the students into assigned groups which could disrupt the ecological validity of the study. However, actions were taken to minimise this limitation using the vocabulary test as a covariate to control the potential influence of pre-existing vocabulary knowledge on the speaking measures.

Second, due to limited access to classroom resources, the present study only included two groups, with both groups essentially experimental groups. It is possible that if a third group were assigned to this study and acted as the control group (i.e. traditional method), the present study would have gained a better understanding of the influence of the different forms of task-related instruction on speaking and students' level of motivation.

Third, the speaking tests were conducted individually and took nearly two weeks for both the pre- and the post-test stage. As mentioned in Chapter 3 (see section 3.4), students were from various engineering backgrounds (e.g. chemical, civil, mechanical etc.), thus they had different scheduling times. During the pre-test week, some students had their speaking test during the first week and some would have it in their second. Therefore the gap between the pre-test and the post-test might differ from one student to another. It would have been preferable to conduct the speaking test within a teaching session and to test all students in each stage simultaneously, or at least within a narrower window of time. In addition, a delayed post-test was not possible due to the time constraints, including the amount of time taken to conduct the pre- and post-

speaking tests. A delayed post-test would have been beneficial to see whether any significant improvement in students' speaking performance and motivation had persisted.

Lastly, the intervention was a mixture of various grammar and lexical items. Previous studies have included one linguistic feature or focused on a specific target structure, such as 15 Spanish nouns (de la Fuente, 2006), relative clauses (Mochizuki & Ortega, 2008), English articles (Ahmadian, 2012), German locative prepositions (Van de Guchte et al., 2017) and past passive verb forms (Ellis et al., 2019; S. Li et al., 2016). However, this was not possible in the present study as the interventions needed to follow the course syllabus in which various grammar structures and vocabulary items are integrated. It is possible that if students had been taught only one specific target structure throughout the intervention, they might have acquired it effectively, especially in the case of grammatical structures.

7.5 Contributions and Implications of the Study

The present study contributes to knowledge in the field and extends the findings of previous research, discussed as follows:

The present study provides empirical evidence to support the model of L2 speech production (Kormos, 2006), which was used in the present study. As mentioned in Chapter Two, the conceptualisation process takes the same form in both L1 and L2. However, the formulation process in L2 differs from the L1 as it requires serial processing, which places higher demands on the speaker including effortful syntax building (Skehan, 2014, 2015). In addition, the present study agrees with the Limited Attention Hypothesis (Skehan, 2009, 2015) that asserts students' resources are limited. The findings confirm that students were not able to balance both complexity and accuracy at the same time during the speaking tests as their attention and capacity were weighted towards one or other of them. These findings support previous studies that conclude that it is unlikely for complexity and accuracy to be elevated simultaneously

(Ahmadian, 2012; Van de Guchte et al., 2017). Besides that, noticing, central to the Noticing Hypothesis (Schmidt, 2010), is considered essential for language acquisition, and is reflected in the study's findings. That is, while the ST group failed to produce accurate language during the formulation process, the EIST group engaged in more reformulation. Even though students in the EIST group did not make significant improvements for all the accuracy measures, that reformulation suggests that they benefitted from noticing and having an awareness of the target linguistic features, as they noticed the incorrect usage of the grammatical items and made the effort to correct those errors.

The study confirms previous other studies in its findings and adds more empirical evidence. For example, the TSLT approach had a positive impact on students' overall vocabulary outcomes and retention of L2 words, but the TBLT approach was more effective in encouraging students to use their own linguistic resources which resulted in a more complex and sophisticated language. These findings are in line with previous studies which suggested that TSLT had positive effects on students' lexical outcomes (de la Fuente, 2006; Ellis et al., 2019; S. Li et al., 2016). Moreover, the TBLT approach made students more fluent in their speech with the reduction of pauses and repairs over time. These findings are in harmony with a recent study which demonstrated that TBLT had a similar effect (Ellis et al., 2019).

However, the study also made new contributions to the existing literature. For example, the previous studies on TBLT were conducted using only quantitative analysis (Ahmadian, 2012; de la Fuente, 2006; Ellis et al., 2019; S. Li et al., 2016; Mochizuki & Ortega, 2008; Van de Guchte et al., 2017). The present study used both mixed-method approach. The quantitative analyses were used to examine the effects of two different forms of language teaching on students' complexity, accuracy and fluency over time. Meanwhile, the qualitative analyses in the present study were adopted to identify the types of vocabulary, forms, errors and disfluency produced by the students before and after the intervention, providing much fuller insights than

previous, solely quantitative studies have been able to do. In addition, the present study highlighted the importance of lemmatisation which was not conducted in the previous studies. This procedure was considered crucial as inflected forms would mask students' lexical knowledge and would affect counting lexical richness in the students' speaking data. Therefore, the qualitative results and the lemmatisation procedure present a new and thorough perspective on students' overall language outcomes.

In regard to motivation, through the interview analysis, it was found that students in the ESL context were more instrumentally motivated after the intervention. These findings corroborate a previous study that found ESL students were more instrumentally motivated due to extrinsic factors (Hong & Ganapathy, 2017). In addition, the findings revealed that the intervention was able to develop students' self-efficacy. This is in line with the previous research that suggested that self-efficacy requires a certain period of time to be developed with sufficient learning experiences and positive role models (Mohamed Khatib & Maarof, 2015). Lastly, situation-based variables (e.g. teacher's role, instruction used, and influence of other ethnic groups in task completion) had a major impact on students' overall motivation towards learning and willingness to communicate in English. These findings are in harmony with previous research which found that explicit guidance from the teacher, different cultural backgrounds and willingness to communicate based on specific situation enhance students' motivation towards language learning (Abdul Razak et al., 2018; de Smedt et al., 2018; S.-J. Kang, 2005).

7.6 Recommendations for Future Research

In light of the above limitations, three areas of possible future research were identified as follows:

First, it would be beneficial to randomly assign students according to their language proficiency (e.g. low intermediate, intermediate and advanced) or different undergraduate major. This

would give the researcher a clear view of whether students from different language levels and backgrounds acquire the grammatical structures and produce lexical items differently.

Second, employing explicit instruction within tasks in different conditions. Previous studies have employed explicit instruction in different variation such as the length of the explicit instruction (e.g. two, five, 10 minutes or 15 minutes), stages where the explicit instruction was employed (e.g. pre-task, during-task or post-task) and using different means to deliver explicit instruction (e.g. handout, exercise, PowerPoint etc.). As for this study, explicit instruction was given for 10 minutes using PowerPoint slides and followed by short grammar/vocabulary exercises. It would be useful to explore explicit instruction in different conditions such as 1) time length of the explicit instruction (e.g. two minutes, five minutes or more than 10 minutes); 2) stages where the explicit instruction is employed (e.g. pre-task, during-task and post-task) and 3) the means by which explicit instruction is presented (e.g. verbalisation, handouts, language exercises, PowerPoint, audio-visual media). Therefore, these conditions could give teachers a deeper understanding of how they influence students' language learning outcomes and motivation to learn the language.

Finally, it would be useful to conduct a longitudinal study to observe the effect of different language teaching methods within a specific curricular context, hence the developmental aspect of grammar and vocabulary acquisition within different teaching approaches can be measured. It is possible that different results would be achieved in a longitudinal study where the gradual effect of explicit instruction is more likely to be evident.

7.7 Summary

The chapter summarised the design of the study, instruments used in the study and the results of the quantitative and the qualitative data analyses. In addition, the chapter also discussed the limitations of the present study and its contribution to empirical evidence in researching explicit

instruction within tasks. Finally, recommendations were given for future research into the effect of task-based language teaching on students' speaking performance and motivation.

REFERENCES

Abdul Rahman, N. H. (2014). *From curriculum reform to classroom practice: An evaluation of the English primary curriculum in Malaysia.* (Doctoral thesis), University of York, York.

Abdul Razak, F. N., Nimehchisalem, V., & Abdullah, A. N. (2018). The relationship between ethnic group affiliation (EGA) and willingness to communicate (WTC) in English among undergraduates in a public university in Malaysia. *International Journal of Applied Linguistics and English Literature, 7*(6), 207-213. doi:10.7575/aiac.ijalel.v.7n.6p.207

Afsahi, S. E., & Lotfi, A. R. (2016). The effect of explicit instruction through combined input-output tasks on the acquisition of indirect reported speech in English. *International Journal of English Linguistics, 6*(7), 59-79. doi:10.5539/ijel.v6n7p59

Ahmadian, M. J. (2012). The effects of guided careful online planning on complexity, accuracy and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research, 16*(1), 129-149. doi:10.1177/1362168811425433

Ahour, T., & Shemshadsara, Z. G. (2015). The impact of focused and unfocused tasks on university students' grammatical achievement. *Theory and Practice in Language Studies, 5*(1), 124-133. doi:10.17507/tpls.0501.17

Albino, G. (2017). Improving speaking fluency in a task-based language teaching appraoch: The case of EFL learners at PUNIV-Cazenga. *SAGE Open, 7*(2), 1-11. doi:10.1177/2158244017691077

Archer, A. L., & Hughes, C. A. (2011). *Explicit instructions: Effective and efficient teaching*. New York: Guilford Press.

Awwad, A., Tavakoli, P., & Wright, C. (2017). "I think that's what he's doing": Effects pf intentional reasoning on second language (L2) speech performance. *System, 67*, 158-169. doi:10.1016/j.system.2017.05.002

Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Bailey, K. M. (2006). Issues in teaching speaking skills to adult ESOL learners. *Review of Adult Learning and Literacy, 6*, 113-164.

Bandura, A. (1989). Human agency in social cognitive theory. *The American Psychologist, 44*(9), 1175-1184.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, USA: W. H. Freeman and Company.

Bandura, A. (1999). Social cognitive theory: An agentic perspective. *Asian Journal of Social Psychology, 2*(1), 21-41. doi:10.1111/1467-839X.00024

Boston, J. (2010). Pre-task syntactic priming and focused task design. *ELT Journal, 64*(2), 165-174. doi:10.1093/elt/ccp033

Brophy, J. (2004). *Motivating students to learn* (2nd ed.). Mahwah, New Jersey: Lawrenc Erlbaum.

Brown, H. D., & Lee, H. (2015). *Teaching by principle: An interactive approach to language pedagogy* (4th ed.). New York, USA: Pearson.

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford, UK: Oxford University Press.

Bui, G., Skehan, P., & Wang, Z. (2018). Task condition effects on advanced-level foreign language performance. In P. A. Malovrh & A. G. Benati (Eds.), *The handbook of advanced proficiency in second language acquisition*. London: John Wiley & Sons.

Bulte, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency*. Amsterdam, Netherlands: John Benjamins Publishin.

Bulte, B., Housen, A., Pierrard, M., & Van Daele, S. (2008). Investigating lexical proficiency development over time: The case of Dutch-speaking learners of French in Brussels. *French Language Studies, 18*, 277-298. doi:10.1017/S0959269508003451

Butler, Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. *Annual Review of Applied Linguistics, 31*, 36-57. doi:10.1017/S0267190511000122

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford, UK: Oxford University Press.

Bygate, M. (2016). Sources, developments and directions of task-based language teaching. *The Language Learning Journal, 44*(4), 381-400. doi:10.1080/09571736.2015.1039566

Chan, S. H., & Tan, H. (2006). English for mathematics and science: Current Malaysian language-in-education policies and practices. *Language and Education, 20*(4), 306-321. doi:dx.doi.org/10.2167/le631.0

Che Musa, N., Koo, Y. L., & Azman, H. (2012). Exploring English language learning and teaching in Malaysia. *Gema Online: Journal of Language Studies, 12*(1), 35-51.

Cheng, H.-F., & Dörnyei, Z. (2007). The use of motivational strategies in language instruction: The case of efl teaching in Taiwan. *Innovation in Language Learning and Teaching, 1*(1), 153-174. doi:10.2167/illt048.0

CHILDES. Child Language Data Exchange System. Retrieved December 15, 2017, from Carnegie Mellon University http://childes.psy.cmu.edu

Choo, J. P. L., & Too, W. K. (2012). Teachers' perception in using task-based instruction (TBI) for the teaching of grammar. *Journal of Interdisciplinary Research in Education, 2*(1), 47-64.

Clément, R. (1980). Ethnicity, contact and communicative competence in a second language. In H. Giles, P. Robinson, & P. Smith (Eds.), *Language: Social psychological perspectives* (pp. 417-154). Oxford: Pergamon.

Clément, R., Baker, S., & MacIntyre, P. D. (2003). Willingness to communicate in a second language: The effects of context, norms, and vitality. *Journal of Language and Social Pscyhology, 22*(2), 190-209. doi:10.1177/0261927X03252758

Clément, R., Dörnyei, Z., & Noels, K. (1994). Motivation, self-confidence and group cohesion in the foreign language classroom. *Language Learning, 44*(3), 417-448. doi:10.1111/j.1467-1770.1994.tb01113.x

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). London, UK: Routledge.

Courtney, L. M. (2014). *Moving from primary to secondary education: An investigation into the effect of primary to secondary transition on motivation for language learning and foreign language proficiency.* (PhD Thesis), University of Southampton, Southampton, England.

Cox, T. L. (2013). *Investigating prompt difficulty in an automatically scored speaking performance assessment* (PhD Thesis), Brigham Young University, London.

Creswell, J. W. (2013a). *Qualitative inquiry and research design: Choosing among five traditions*. Thousan Oaks, California: Sage Publication.

Creswell, J. W. (2013b). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Los Angeles, USA: Sage.

Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, USA: Pearson.

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research*. Thousands Oaks: SAGE Publications.

Crookes, G., & Schmidt, R. (1991). Motivation: Reopening the research agenda. *Language Learning, 41*(4), 469-512. doi:10.1111/j.1467-1770.1991.tb00690.x

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. London, UK: Sage.

Cziser, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. In M. Nikolov (Ed.), *Early learning of modern foreign languages : processes and outcomes* (pp. 98-117). Bristol: Multilingual Matters.

Dalili, M. V. (2011). On the integration of form and meaning in English language teaching: An overview of current pedagogical options. *Procedia: Social and Behavioural Sciences, 15*, 2117-2121. doi:10.1016/j.sbspro.2011.04.064

Daller, M. (2010). *Guiraud's index of lexical richness.* Paper presented at the Applied Linguistics, Global and Local, University of Aberdeen.

Daller, M., & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students: A comparison of different measures. In M. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 150-164). Cambridge, UK: Cambridge University Press.

Darmi, R., & Albion, P. (2013). *English language in the Malaysian education system: Its existence and implications*. Paper presented at the 3rd Malaysian Postgraduate Conference, Sydney, Australia.

Darus, S. (2010). The current situation and issues of the teaching of English in Malaysia. *Ritsumeikan Studies in Language and Culture, 22*(1), 19-27.

David, M. K., Milde, A., & Coluzzi, P. (2009). Language policies - impact on language maintenance and teching: Focus on Malaysia, Singapore, Brunei and the Philippines. *The Linguistics Journal*, 155-191.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics, 13*, 1-24.

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46*(3), 529-555. doi:10.1111/j.1467-1770.1996.tb01246.x

de Jong, N., & Bosker, H. R. (2013). *Choosing a threshold for silent pauses to measure second language fluency.* Paper presented at the The 6th Workshop on Disfluency in Spontaneous Speech, Stockholm, Sweden.

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533-568. doi:10.1111/j.1467-9922.2010.00620.x

de la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research, 10*(3), 263-295. doi:10.1191/1362168806lr196oa

de Smedt, F., Graham, S., & van Keer, H. (2018). The bright and dark side of writing motivation: Effects of explicit instruction and peer assistance. *The Journal of Educational Research, 1*(16). doi:10.1080/00220671.2018.1461598

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*(3), 379-410.

DeKeyser, R. (1996). Exploring automatization processes. *TESOL Quarterly, 30*(2), 349-357. doi:10.2307/3588151

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition*. Cambridge, UK: Cambridge University Press.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97-113). New Jersey: Erlbaum.

DeKeyser, R., & Criado, R. (2013). Automatization, skill acquisition and practice in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford, UK: Blackwell Publishing.

Department of Statistics Malaysia. (2018). Current population estimates, Malaysia, 2017-2018. Retrieved from https://bit.ly/2Qe0ykE

Dincer, A., & Yesilyurt, S. (2017). Motivation to speak English: A self-determination theory perspective. *PASAA, 53*, 1-25.

Dornyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford, UK: Oxford University Press.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal, 78*(3), 273-284.

Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (Ed.), *Individual differences and instructed language learning* (Vol. 2, pp. 137-158). Amsterdam, Netherlands: John Benjamins Publishing.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. New Jersey: Lawrence Erlbaum.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research, 4*(3), 275-300. doi:10.1177/136216880000400305

Dornyei, Z., & Taguchi, T. (2010). *Questionnaire in second language research: Construction, administration and processing* (2nd ed.). London, UK: Routledge.

Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Harlow, England: Pearson.

Doughty, C., & Pica, T. (1986). Information gap' tasks: Do they facilitate second language acquisition? *TESOL Quarterly, 20*(2), 305-325. doi:10.2307/3586546

Doughty, C., & Williams, J. (1998a). Issues and terminology. In C. Doughty & J. Williams (Eds.), *Focus on from in classroom second language acquisition* (pp. 1-12). Cambridge: Cambridge University Press.

Doughty, C., & Williams, J. (1998b). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197-262). Cambridge: Cambridge University Press.

Elder, C., & Iwashita, N. (2005). Planning for test performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (Vol. 11, pp. 219-238). Amsterdam: John Benjamins Publishing.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27,* 164-194.

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly, 27*(1), 91-113. doi:10.2307/3586953

Ellis, R. (2003). *Task-based language teaching and learning*. Oxford, UK: Oxford University Press.

Ellis, R. (2005a). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-172. doi:10.1017/S0272263105050096

Ellis, R. (2005b). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (Vol. 11). Amsterdam, Netherlands: John Benjamin Publishing.

Ellis, R. (2006). Current issue in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*(1), 83-107. doi:10.2307/40264512

Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit*

knowledge in second language learning, testing and teaching. Bristol: Multilingual Matters.

Ellis, R. (2009b). *The methodology of task-based teaching.* Paper presented at the The Asian EFL Journal Cebu Conference Proceedings.

Ellis, R. (2009c). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics, 19*(3), 221-246.

Ellis, R. (2016). Focus on form: A critical review. *Language Teaching Research, 20*(3), 405-428. doi:10.1177/1362168816628627

Ellis, R. (2017). Moving task-based language teaching forward. *Language Teaching, 50*(4), 507-526.

Ellis, R. (2018). *Reflections on task-based language teaching*. Bristol, UK: Multilingual Matters.

Ellis, R. (2019). Towards a modular language curriculum for using task. *Language Teaching Research, 23*(4), 454. doi:10.1177/1362168818765315

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford, UK: Oxford University Press.

Ellis, R., Li, S., & Zhu, Y. (2019). The effects of pre-task explicit instruction on the performance of focused task. *System, 80*, 38-47. doi:10.1016/j.system.2018.10.004

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. London: Routledge.

Fields, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). London: Sage.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*(3), 299-323. doi:10.1017/S0272263100015047

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375. doi:10.1093/applin/21.3.354

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics, 36*, 98-116. doi:10.1017/S0267190515000082

Gardner, R. C. (1985). *Social pyschology and second language learning: The role of attitude and motivation*: Edward Arnold.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Gardner, R. C., & MacIntyre, P. D. (1992). A student's contributions to second-language learning. Part I: Cognitive variables. *Language Teaching*, *25*(4), 211-220. doi:10.1017/S026144480000700X

Gay, L. R., Mills, G. E., & Airasian, P. W. (2011). *Educational research: Competencies for analysis and applications* (10th ed.). USA: Pearson Education.

Gray, D. E. (2014). *Doing research in the real wold* (3rd ed.). California, USA: Sage Publication.

Greene, J., Benjamin, L., & Goodyear, L. (2001). The merits of mixing methods for evaluation. *Evaluation*, *7*(1), 25-44.

Guba, E. (1990). *The paradigm dialog*. Newbury Park, California: SAGE Publication.

Guillot, M.-N. (1999). *Fluency and its teaching*. Clevedon: Multilingual Matters.

Guz, E. (2015). Establishing the fluency gap between native and non-native speech. *Research in Language, 13*(3), 230-427. doi:10.1515/rela-2015-0021

Hall, S. J. (2015). Gaining acceptance of task-based teaching during Malaysian rural in-service teacher training. In M. Thomas & H. Reinders (Eds.), *Contemporary task-based language teaching in Asia* (pp. 156-169). London: Bloomsbury.

Han, Z., & Finneran, R. (2013). Re-engaging the interface debate: Strong, weak, none, or all? *International Journal of Applied Linguistics, 24*(3), 1-20. doi:10.1111/ijal.12034

Harper, M., & Cole, P. (2012). Member checking: Can benefits be gained similar to group therapy? *The Qualitative Report, 17*(2), 1-8.

Heng, C. S., Abdullah, A. N., & Yusof, N. (2012). Investigating the construct of anxiety in relation to speaking skills among ESL tertiary learners. *3L: The Southeast Asian Journal of English Language Studies, 18*(3), 155-166.

Hong, Y. C., & Ganapathy, M. (2017). To investigate ESL students' intrumental and integrative motivation towards English language leanring in a Chinese school in Penang: Case study. *English Language Teaching, 10*(9), 17-35. doi:10.5539/elt.v10n9p17

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473. doi:10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (Vol. 32). Amsterdam: John Benjamins Publishing.

Housen, A., & Pierrard, M. (2005). Investigating instructed second language acquisition. In A. Housen & M. Pierrard (Eds.), *Investigating in instructed second language acquisition*. Berlin: Mouton de Gruyter.

Hulstijin, J. (2005). Theoretical and empirical issues in the study of implicit and explciti second language learning. *Studies in Second Language Acquisition, 27*(2), 129-140. doi:10.1017/S0272263105050084

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*(3), 193-223. doi:10.1191/0267658302sr207oa

Hulstijn, J. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (Vol. 48, pp. 25-46). Amsterdam: John Benjamins Publishing.

Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English, 43*(7), 732-739.

Idrus, H., & Salleh, R. (2008). Perceived self-efficacy of Malaysian ESL engineering and technology students on their speaking ability and its pedagogical implications. *The English Teacher, 37*, 61-75.

Inoue, C. (2016). A comparative study of the variables used to measure syntactic complexity and accuracy in task-based research. *The Language Learning Journal, 44*(4), 487-505. doi:10.1080/09571736.2015.1130079

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3-20. doi:10.1177/1525822X05282260

Iwanaka, T., & Takatsuka, S. (2007). Influences of attention and noticing on second language acquisition. *Languages and Literatures*, 21-30.

Jarvis, S. (2002). Short text, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57-84.

Ji, Y., & Pham, T. (2018). Implementing task-based language teaching (TBLT) to teach grammar in English classes in China: Using design-based research to explore challenges and strategies. *Innovation in Language Learning and Teaching*, 1-14. doi:10.1080/17501229.2018.1545021

Julkunen, K. (2001). Situation and task-specific motivation in foreign language learning. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acqusition* (pp. 29-42). Hawaii, USA: University of Hawaii Press.

Kahneman, D. (1973). *Attention and effort*. New Jersey, USA: Prentice Hall.

Kang, O., Rubin, D., & Pickering, L. (2010). Sugrasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal, 94*(4), 554-566. doi:10.1111/j.1540-4781.2010.01091.x

Kang, S.-J. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System, 33*(2), 277-292. doi:10.1016/j.system.2004.10.004

Kaur, N., & Nordin, R. (2006). A case for reconstruction of pedagogy of Malaysian university English test (MUET) through thematic units of intruction. *Journal of Institutional Research South East Asia, 4*(1), 5-16.

Kayad, F. G. (2015). *Literary literacy for professional preparation or pre-service and conversion English teachers in Malaysia.* (Doctoral thesis), The University of Western Australia, Australia.

Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction, 1*(1), 60-69.

Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System, 40*, 554-564. doi:10.1016/j.system.2012.10.012

Kormos, J. (2006). *Speech production and second language acquisition*. New Jersey, USA: Lawrence Erlbaum Associates.

Kormos, J. (2011). Speech production and the cognition hypothesis. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 39-60). Amsterdam, Netherlands: John Benjamins Pubslihing.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Prentice Hall.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Kremmel, B., & Schmidt, N. (2018). Vocabulary levels test. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching*: John Wiley.

Krüsemann, H. (2018). *Language learning motivation and the discursive representations of German, the Germans, and Germany in UK school settings and the press* (PhD Thesis), University of Reading, Reading.

Kumaravadivelu, B. (2006). TESOL methods: Changing tracks, challenging trends. *TESOL Quarterly*, *40*(1), 59-81.

Laufer, B. (2006). Comparing focus on form and focus on forms in second language vcabulary learning. *The Canadian Modern Language Review, 63*(1), 149-166.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. London: Sage.

Lennon, P. (1999). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417. doi:doi.org/10.1111/j.1467-1770.1990.tb00669.x

Leow, R. P. (2015). The role of attention and (un)awareness in SLA: Conceptual replication of N. C. Ellis & Sagarra (2010a) and Leung & Williams (2012). *Language Teaching, 48*(1), 117-129. doi:10.1017/S0261444814000275

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Massachusetts, USA: MIT Press.

Li, L., Chen, J., & Sun, L. (2015). The effects of different lengths of pretask planning time on L2 learners' oral test performance. *TESOL Quarterly, 49*(1), 38-66.

Li, S., Ellis, R., & Zhu, Y. (2016). Task-based versus task-supported language instruction: An experimental study. *Annual Review of Applied Linguistics, 36*, 205-229. doi:0.1017/S0267190515000069

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot & T. Huebner (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins Publishing.

Long, M. (2000). Focus on form in task-based language teaching. In R. D. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton* (Vol. 179-192). Amsterdam: John Benjamins Publishing.

Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell.

Long, M. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics, 36,* 5-33. doi:10.1017/S0267190515000057

Long, M., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambrdige: Cambridge University Press.

MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal, 84*(4), 545-562. doi:10.1111/j.1540-4781.1998.tb05543.x

Mackey, A. (1999). Input, interaction adn second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*(4), 557. doi:10.1017/S0272263199004027

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405-430. doi:10.1093/applin/ami051

Mackey, A. (2012). *Input, interaction and corrective feedback in L2 learning*. Oxford: Oxford University Press.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, New Jersey: Lawrence Erlbaum.

MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malakolunthu, S., & Rengasamy, N. C. (2012). Education policies and pracices to address cultural diveristy in Malaysia: Issues and challenges. *Prospects, 42*(2), 147-159. doi:10.1007/s11125-012-9227-9

Malicka, A., & Levkina, M. (2012). Measuring task complexity: Does EFL proficiency matter? In S. Ali & C. A. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (Vol. 4, pp. 43-66). Amsterdam, Netherlands: John Benjamins.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing, 19*(1), 85-104. doi:10.1191/0265532202lt221oa

Mansor, N., & Hasan, M. J. (2015). *Communicative approach and the teaching of speaking skill*. Terengganu: Universiti Malaysia Terengganu.

Mat Hassan, M. A., & Talib, R. (2013). *Perception towards SBA implementation among teachers in Malaysian schools*. Paper presented at the 2nd International Seminar on Quality and Affordable Education, Johor, Malaysia.

McCarthy, P. M., & Jarvis, S. (2007). VOCD: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488. doi:10.1177/0265532207080767

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behaviour Research Methods, 42*(2), 381-392. doi:10.3758/BRM.42.2.381

McCroskey, J. C., & Baer, J. E. (1985). *Willingness to communicate: The construct and its measurement*. Paper presented at the The Annual Meeting of the Speech Communication Association, Denver, CO.

McCroskey, J. C., & Richmond, V. P. (1990). Willingness to communicate: Differing cultural perspectives. *Sourthern Communication Journal, 56*(1), 72-77. doi:10.1080/10417949009372817

McDonough, J., Shaw, C., & Masuhara, H. (2013). *Materials and methods in ELT: A teacher's guide* (3rd ed.). West Sussex: Blackwell Publishing.

McGrath, I. (2016). *Material evaluation and design for language teaching*. Edinburgh: Edinburgh University Press.

Meakins, F. (2007). Computerized Language Analysis (CLAN) from The CHILDES Project. *Language Documentation and Conversation, 1*(1), 107-112.

MEC. (2015). *Malaysian university English test (MUET): Regulations, test specifications, test format and sample questions* (3rd ed.). Selangor: Malaysian Examinations Council.

Mechura, M. B. (2016). Lemmatization lists. Retrieved December 27, 2017 http://www.lexiconista.com/datasets/lemmatization/

Mehnert, U. (1998). The Effects of Different Length of Time for Planning on Second Language Performance. *Studies in Second Language Acquisition, 20*(1), 83-108.

Michael, E. (2012). English teaching policy analysis: Flaws and the implications among Malaysian secondary schools students and teachers. *Journal of Humanities and Social Sciences, 1*(1), 62-72.

Michel, M. C., Kuiken, F., & Vedder, I. (2012). Task complexity and interaction: (Combined) effects on task-based performance in Dutch as a second language *EUROSLA yearbook* (Vol. 12, pp. 164-190). Amsterdam: John Benjamins.

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in the beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research, 12*(1), 11-37. doi:10.1177/1362168807084492

MoE. (2013). *Malaysia education blueprint 2013-2025: Preschool to post-secondary education*. Putrajaya, Malaysia: Ministry of Education Malaysia.

MoE. (2014). *Dasar memartabatkan bahasa Malaysia memperkukuh bahasa Inggeris (MBMMBI): Buku penerangan*. Putrajaya, Malaysia: Ministry of Education Malaysia.

MoE. (2015a). *Bahasa Inggeris: Dokumen standard kurikulum dan pentaksiran tingkatan 1*. Putrajaya, Malaysia: Ministry of Education Malaysia.

MoE. (2015b). *Malaysia education blueprint 2015-2025: Higher education*. Putrajaya, Malaysia: Ministry of Education Malaysia.

MoE. (2018). *Bahasa Inggeris sekolah kebangsaan: Dokumen standard kurikulum dan pentaksiran*. Putrajaya, Malaysia: Ministry of Education, Malaysia.

Mohamed Khatib, F. M., & Maarof, N. (2015). Self-efficacy perception of oral communication ability among English as a second language (ESL) technical students. *Procedia: Social and Behavioural Sciences, 204*, 98-104. doi:10.1016/j.sbspro.2015.08.121

Muftah, M., & Rafik-Galea, S. (2015). Language learning motivation among Malaysian pre-university students. *English Language Teaching, 6*(3), 92-103. doi:10.5539/elt.v6n3p92

Muhamad, A. J., Ahamad Shah, M. I., Engku Ibrahim, E. H., Sarudin, I., Abdul Malik, F., & Abdul Ghani, R. (2013). Oral presentation errors of Malaysian students in an English for Academic Purposes (EAP) course. *World Applied Sciences Journal, 21*, 19-97. doi:10.5829/idosi.wasj.2013.21.sltl.2133

Murdock, S. H., & Ellis, R. (1991). *Applied demography: An introduction to basic concepts, methods and data*. Colorado, USA: Westview Press.

Muslim, A. (2013). *An investigation into the oral communication skills component in the English language syllabuses of the Malaysian polytechnics.* (Doctoral thesis), University of Strathclyde, Strathclyde.

Mustafa, Z. (2012). Teachers' encountered challenges in the adoption of task-based language teaching in Malaysian classrooms. *International Journal of Arts and Sciences, 5*(3), 269-279.

Mustapha, G. (2008). A reflection of the revised syllabus translated in textbooks. *Journal of international Managament Studies*, 162-169.

Nakamura, Y. (2009). Rating criteria for the three speaking test format: Monologue, dialogue and multilogue. *Educational Studies, 51*, 133-141.

NamazianDost, I., Bohloulzadeh, G., & Pazhakh, A. (2017). The effect of task-based language teaching on motivation and grammatical achievement of EFL junior high school students. *Advances in Language and Literary Studies, 8*(2), 243-259. doi:10.7575/aiac.alls.v.8n.2p.243

Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. New York: Routledge.

Neff, P. E. (2007). The role of anxiety and motivation in language learner task performance. *Doshisha Studies in Language and Culture, 10*(1), 23-42.

Neuman, W. L. (2006). *Social research methods: Qualitative and quantitative approaches* (6th ed.). London, UK: Pearson.

Newby, P. (2014). *Research methods for education* (2nd ed.). London, UK: Routledge.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 intruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528. doi:10.1111/0023-8333.00136

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578. doi:10.1093/applin/amp044

Nunan, D. (1993). Task-based syllabus design: Selecting, grading and sequencing tasks. In G. Crookes & S. M. Gass (Eds.), *Tasks in a pedagogical context: Integrating theory and practice* (pp. 55-68). Clevedon, UK: Multilingual Matters.

Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia-Pacific region. *TESOL Quarterly, 37*(4), 589-613. doi:10.2307/3588214

Nunan, D. (2004). *Task-based language teaching*. Cambridge, UK: Cambrige University Press.

Ochieng, P. A. (2009). An analysis of the strengths and limitation of qualitative and quantitative research paradigms. *Problems of education in the 21st century, 13*, 13-18.

Ortega, L. (2007). Meaningful L2 practice in foreign language classrooms: A cognitive-interactionist SLA perspective. In R. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 180-207). Cambridge, UK: Cambridge University Press.

Palloti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590-601. doi:10.1093/applin/amp045

Pandian, A. (2002). English language teaching in Malaysia today. *Asia Pacific Journal of Education, 22*(2), 35-52. doi:10.1080/0218879020220205

Pang, F., & Skehan, P. (2014). Self-reported planning behaviour and second language performance in narrative retelling. In P. Skehan (Ed.), *Processing perspective on task performance* (Vol. 5, pp. 97-127). Amsterdam, Netherlands: Jon Benjamins Publishing.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419). London: Academic Press.

Pennycook, A. (2017). *The cultural politics of English as an international language*. London: Routledge.

Philp, J., & Cao, Y. (2006). Interactional context and willingness to communicate: A comparison of behavior in whole class, group and dyadic interaction. *System, 34*(4), 480-493. doi:10.1016/j.system.2006.05.002

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition, 11*(1), 63-90. doi:10.1017/S027226310000783X

Pillai, S., Khan, M. J., Ibrahim, I. S., & Raphael, S. (2012). Enhancing employability through industrial training in the Malaysia context. *Higher Education, 63*(2), 187-204. doi:10.1007/s10734-011-9430-2

Plonsky, L., & Oswald, F. L. (2014). How big is "big"?: Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912. doi:10.1111/lang.12079

Pollard, A. J. (2012). *Speech production, dual-process theory and the attentive addressee.* (Doctoral thesis), University College London, London, UK.

Poulisse, N. (1997). Language productions in bilinguals. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*. New Jersey, USA: Lawrence Erlbaium Associates.

Punch, K. F. (2009). *Introduction to research methods in education*. California, USA: Sage Publication.

Qian, L. (2014). Get it right in the end: The effects of post-task transcribing on learners' oral performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (Vol. 5, pp. 129-154). Amsterdam, Netherlands: John Benjamins.

Qiu, X. (2019). Functions of oral monologic tasks: Effects of topic familiarity on L2 speaking performance. *Language Teaching Research*, 1-20. doi:10.1177/1362168819829021

Queiros, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods *European Journal of Education Studies, 3*(9), 369-387.

Rashid, R. A., Abdul Rahman, S. B., & Yunus, K. (2017). Reforms in the policy of English language teaching in Malaysia. *Policy Futures in Education, 15*(1), 100-112. doi:10.1177/1478210316679069

Reinbold, S. (2013). Using the ADDIE model in designing library instruction. *Medical Reference Services Quarterly, 32*(3), 244-256. doi:10.1080/02763869.2013.806859

Rethinasamy, S., & Chuah, K. M. (2011). The Malaysian university English test (MUET) and its use for placement purposes: A predictive validity study. *Electronic Journal of Foreign Language Teaching, 2*, 234-245.

Robinson, P. (1995). Attention, memory and the noticing hypothesis. *Language Learning, 45*(2), 283-331. doi:10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (2001). Task complexity, task difficulty and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27-57.

Robinson, P. (2003). Attention and memoring during SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Cornwall: Blackwell Publishing.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-27). Clevedon, Uk: Multilingual Matters.

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics, 45*(3), 161-176. doi:10.1515/IRAL.2007.007

Rogers, J., Revesz, A., & Rebuschat, P. (2016). Implicit and explicit knowledge of inflectional morphology. *Applied Psycholinguistics, 37*(4), 781-812. doi:10.1017/S0142716415000247

Samuda, V. (2009). Guiding relationships between form and meaning during task performance. In K. V. Den Branden, M. Bygate, & J. M. Norris (Eds.), *Task-based language teaching: A reader* (pp. 379-400). Amsterdam: John Benjamins Publishing.

Sandelowski, M. (2001). Real qualitaive researchers do not count: The use of numbers in qualitative research. *Research in Nursing and Health, 24*, 230-240. doi:10.1002/nur.1025

Sato, R. (2010). Reconsidering the effectiveness and suitability of PPP and TBLT in the Japanese EFL classroom. *JALT Journal, 32*(2), 189-201.

Schmidt, R. (1990). The role of consiousness in second language learning. *Applied Linguistics, 11*(2), 129-158. doi:10.1093/applin/11.2.129

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Reviews, 11*, 237-326.

Schmidt, R. (2010). Attention, awareness and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC* (2-4 December ed., pp. 721-737). Singapore: Centre for Language Studies, National University of Singapore.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing, 18*(1), 55-88. doi:10.1177/026553220101800103

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In K. R. Wentzel & D. B. Meile (Eds.), *Handbook of motivation at school* (pp. 34-54). London: Routledge.

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 382-408). Oxford, UK: Blackwell.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Selvaraj, B. (2010). English language teaching (ELT) curriculum reforms in Malaysia. *Voice of Academia, 5*(1), 51-60.

Shehadeh, A. (2005). Task-based language learning and teaching: Theories and application. In C. Edwards & J. Willis (Eds.), *Teachers exploring tasks in English language teaching*. Houndmills: Palgrave Macmillan.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*(1), 38-62. doi:10.1093/applin/17.1.38

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integratting complexity accuracy fluency and lexis. *Applied Linguistics, 30*(4), 510-532. doi:10.1093/applin/amp047

Skehan, P. (2014). The context for researching of processing perspective on task performance. In P. Skehan (Ed.), *Processing perspective on task performance* (Vol. 5, pp. 1-26). Amsterdam, Netherlands: John Benjamins Publishing.

Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (Vol. 8, pp. 123-155). Amsterdam, Netherlands: John Benjamins Publishing.

Skehan, P. (2018). *Second language task-based performance: Theory, research and assessment*. New York: Routledge.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*(1), 93-120. doi:10.1111/1467-9922.00071

Skehan, P., & Foster, P. (2007). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language insturction* (3rd ed., pp. 183-205). Cambridge: University of Cambridge Press.

Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32, pp. 199-220). Amsterdam, Netherlands: John Benjamins Pubslishing.

Smith, J. (1989). *The nature of social and education inquiry: Empiricism versus interpretation*. Norwoord, New Jersey: Ablex.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and types of language feature: A meta-analysis. *Language Learning, 60*(2), 263-308. doi:10.1111/j.1467-9922.2010.00562.x

Storch, N., & Wigglesworth, G. (2007). Writing tasks: The effects of collaboration. In M. d. P. G. Mayo (Ed.), *Investigating tasks in formal language learning*. New York, USA: Multilingual Matters.

Sulaiman, T., Mohd Ayub, A. F., & Sulaiman, S. (2015). Curriculum change in English language curriculum advocates higher order thinking skills and standards-based assessments in Malaysian primary schools. *Mediterranean Journal of Social Sciences, 6*(2).

Sulan, N., Abdul Majid, N., Kunjambu, B., Sabun, S., Osman, N., & Ismail, R. (2014). *UHL 2400: Fundamentals of English language*. Pahang, Malaysia: Penerbit Universiti Malaysia Pahang.

Talmy, S. (2010). Qualitative interviews in applied linguistics: From research instrument to social practice. *Annual Review of Applied Linguistics, 30*, 128-149. doi:10.1017/S0267190510000085

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*(1), 71-79. doi:10.1093/elt/ccq020

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics and Language Teaching, 54*(2), 1-14. doi:10.1515/iral20169994

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439-473. doi:10.1111/j.1467-9922.2008.00446.x

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language*. Netherlands: John Benjamins Publishing.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, California: SAGE Publication.

Teng, H. C. (2007). *A study of task types of L2 speaking assessment*. Paper presented at the Annual Meeting of the International Society for Language Studies, Honolulu, Hawaii.

Tharmalingam, S. (2012). *Language policy changes in Malaysia: Progressive or regressive?* Paper presented at the 2nd International Conference on Linguistics, Literature and Culture, Penang, Malaysia.

Thirusanku, J., & Md Yunus, M. (2012). The many faces of Malaysian English. *International Scholarly Research Network*, 1-14. doi:10.5402/2012/138928

Thornbury, S. (2005). *How to teach speaking*. Essex: Longman.

Tidball, F., & Treffers-Daller, J. (2007). Exploring measures of vocabulary richness in semi-spontaneous French speech: A quest for the Holy Grail? In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 133-148). Cambridge, UK: Cambridge University Press.

Tollefson, J. W., & Tsui, A. B. M. (2003). The centrality of medium-of-instruction policy in sociopolitical processes. In J. W. Tollefson & A. B. M. Tsui (Eds.), *Medium of instruction policies: Which agenda? Whose agenda?* (pp. 1-18). New York: Routledge.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16,* 183-203.

Too, W. K. (2017). English language teaching and policies at the tertiary level in Malaysia. In E. S. Park & B. Spolsky (Eds.), *English education at the tertiary level in Asia: From policy to practice* (pp. 109-129). London: Routledge.

Treffers-Daller, J. (2009). Language dominance and lexical diversity: How bilinguals and L2 learners differ in their knowledge and use of French lexical and functional items. In B. Richards, H. M. Daller, D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition:The Interface between theory and application*. Basingstoke, England: Palgrace MacMillan.

Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of french: An exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Jarvis & M. Daller (Eds.), *Vocabulary Knowledge: Human Ratings and Automated Measures* (pp. 79-104). Amsterdam, Netherlands: John Benjamins.

Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics, 39*(3), 302–327. doi:10.1093/applin/amw009

Tunku Mohtar, T. M., Swaran Singh, C. K., Abdullah, N. Y., & Mat, M. (2015). Developing the speaking skill among ESL elementary learners. *The Asian Journal of English Language and Pedagogy, 3,* 161-180.

Ur, P. (2013). Language-teaching method revisited. *ELT Journal, 67*(4), 468-474. doi:10.1093/elt/cct041

Van de Guchte, M., Rijlaarsdam, G., Braaksma, M., & Bimmel, P. (2017). Focus on language versus content in pre-task: Effect of guided peer-video model observations on task performance. *Language Teaching Research*, 1-20. doi:10.1177/1362168817735543

van Hout, R., & Vermeer, A. (2007). Comparing measures of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 93-115). Cambridge, UK: Cambridge University Press.

van Lier, L. (2004). *The ecology and semiotics of language learning*. Boston: Kluwer Academic.

VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory research and commentary*. Mahwah, New Jersey: Lawrence Erlbaum.

Varcelotti, M. L. (2012). *Complexity, accuracy and fluency as properties of language performance: The development of the multiple subsystems over time and in relation to each other.* (Doctoral thesis), University of Pittsburgh, USA.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65-83.

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 Children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 173-189). Philadelphia, USA: John Benjamins.

Vethamani, M. E. (2007). The ebb and flow of English language education in Malaysia. In M. E. Vethamani & R. Perumal (Eds.), *Teaching English in Malaysia* (pp. 1-10). Petaling Jaya, Malaysia: Sasbadi.

Warren, P. (2013). *Introducing psycholinguistics*. Cambridge, UK: Cambridge University Press.

Wellington, J. (2015). *Educational research : Contemporary issues and practical approaches*. London, England: Bloomsbury.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*(1), 85-106. doi:10.1177/026553229701400105

Williams, M. (1994). Motivation in foreign and second language learning: An interactive perspective. *Educational and Child Psychology, 11*(2), 77-84.

Willis, J. (1996). *A flexible framework for task-based learning*. Essex, UK: Longman.

Willis, J. W. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. Thousand Oaks, California: SAGE Publication.

Wright, H. H., Silverman, S., & Newhoff, M. (2003). Measures of lexical diveristy in aphasia. *Aphasiology, 17*(5), 443-452.

Yang, L. F., & Ahmad Ishak, M. S. (2012). Framing controversy over language policy in Malaysia: The coverage of PPSMI reversal (teaching of mathematics and science in English) by Malaysian newspapers. *Asian Journal of Communication, 22*(5), 449-473. doi:10.1080/01292986.2012.701312

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal, 86*(1), 54-66. doi:10.1111/1540-4781.00136

Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning, 54*(1), 119-152. doi:10.1111/j.1467-9922.2004.00250.x

Zainol Abidin, M. J., Pour-Mohammadi, M., & Alzwari, H. (2012). EFL students' attitudes towards learning English language: The case of Libyan secondary school students. *Asian Social Science, 8*(2), 119-134. doi:10.5539/ass.v8n2p119

Zhang, S. (2009). The role of input, interaction and output in the development of oral fluency. *English Language Teaching, 2*(4), 91-100.

Zúñiga, E. C. (2016). Implementing task-based language teaching to integrate language skills in an EFL program at a Colombian university. *Issues in Teachers' Professional Development, 18*(2), 13. doi:10.15446/profile.v18n2.49754

APPENDICES

Appendix A: Lesson Plan and Activities for Lesson One

Theme	:	Environment
Topic	:	Save our planet!
Grammar	:	Review of nouns and adjectives
Vocabulary	:	Describing the noun (e.g. environment) using adjectives

Task procedure

Pre-task

1. Teacher plays a video from YouTube about plastic pollution in the sea and students (EIST group) were required to focus on the language.

2. Teacher posts up a few discussion questions relating to the topic (e.g. *What is the video about? How do you feel? What would you do?*) to elicit students' interest.
3. Teacher asks the students if they remember nouns and adjectives, and reviews the grammar items based on the theme.
4. Students are required to identify the grammar items (nouns and adjectives) and fill in the blanks with missing vocabulary based on audio by the National Geographic entitled Air Polution 101.

During-task

Group discussion

1. Students are required to form a group of 3 to 4 students.
2. They are required to discuss on how to avoid or stop pollution and make the world a greener place.
3. Students are given a large sketch paper and several marker pens to draw and write their ideas in the most creative way.
4. Each group must come to a consensus on what is the best solution and present their ideas in front of the class.
5. Students are reminded to use the grammar and vocabulary items which were previously taught in the pre-task stage.

Group presentation

1. Within the same group, students are asked to present their ideas in front of the class.
2. While one group presents, other groups are asked to be alert with the use of the three grammar items and correct them if the group presenting uses them wrongly.

Post-test

1. Teacher writes the errors made by the students during the group presentations.
2. Teachers recaps the topic, grammar and vocabulary items used in the classroom.

Appendix B: Vocabulary Levels Test (Schmitt et al., 2001)

Researcher:
Muhammad Yasir Yahya
Email: m.y.yahya@pgr.reading.ac.uk

VOCABULARY LEVELS TEST

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning. Here is an example.

1	business	
2	clock	___ part of a house
3	horse	___ animal with four legs
4	pencil	___ something used for writing
5	shoe	
6	wall	

You answer it in the following way.

l	business	
2	clock	6 part of a house
3	horse	*3* animal with four legs
4	pencil	*4* something used for writing
5	shoe	
6	wall	

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are business, clock, and shoe. If you have no idea about the meaning of a word, do not guess. But if you think you might know the meaning, then you should try to find the answer.

2000 WORD LEVEL

1	birth		
2	dust	___	game
3	operation	___	winning
4	row	___	being born
5	sport		
6	victory		

1	adopt		
2	climb	___	go up
3	examine	___	look at closely
4	pour	___	be on every side
5	satisfy		
6	surround		

1	choice		
2	crop	___	heat
3	flesh	___	meat
4	salary	___	money paid
5	secret		regularly for
6	temperature		doing a job

1	bake		
2	connect	___	join together
3	inquire	___	walk without
4	limit		purpose
5	recognize	___	keep within a
6	wander		certain size

1	cap	___	teaching and
2	education		learning
3	journey	___	numbers to
4	parent		measure with
5	scale	___	going to a far
6	trick		place

1	burst		
2	concern	___	break open
3	deliver	___	make better
4	fold	___	take something to
5	improve		someone
6	urge		

1	attack		
2	charm	___	gold and silver
3	lack	___	pleasing quality
4	pen	___	not having
5	shadow		something
6	treasure		

1	original		
2	private	___	first
3	royal	___	not public
4	slow	___	all added together
5	sorry		
6	total		

1	cream		
2	factory	___	part of milk
3	nail	___	a lot of money
4	pupil	___	person who is
5	sacrifice	___	studying
6	wealth		

1	brave		
2	electric	___	commonly done
3	firm	___	wanting food
4	hungry	___	having no fear
5	local		
6	usual		

3000 WORD LEVEL

1	belt	___	idea	1	betray		
2	climate	___	inner surface of	2	dispose	___	frighten
3	executive		your hand	3	embrace	___	say publicly
4	notion	___	strip of leather	4	injure	___	hurt seriously
5	palm		worn around the	5	proclaim		
6	victim		waist	6	scare		

1	acid			1	encounter		
2	bishop	___	cold feeling	2	illustrate	___	meet
3	chill	___	farm animal	3	inspire	___	beg for help
4	ox	___	organization or	4	plead	___	close completely
5	ridge		framework	5	seal		
6	structure			6	shift		

1	bench			1	assist		
2	charity	___	long seat	2	bother	___	help
3	jar	___	help to the poor	3	condemn	___	cut neatly
4	mate	___	part of a country	4	erect	___	spin around
5	mirror			5	trim		quickly
6	province			6	whirl		

1	boot			1	annual		
2	device	___	army officer	2	concealed	___	wild
3	lieutenant	___	a kind of stone	3	definite	___	clear and certain
4	marble	___	tube through	4	mental	___	happening once a
5	phrase		which blood	5	previous		year
6	vein		flows	6	savage		

1	apartment	___	a place to live	1	dim		
2	candle	___	chance of	2	junior	___	strange
3	draft		something	3	magnificent	___	wonderful
4	horror		happening	4	maternal	___	not clearly lit
5	prospect	___	first rough form	5	odd		
6	timber		of something written	6	weary		

5000 WORD LEVEL

1	balloon	___	bucket
2	federation	___	unusual
3	novelty		interesting thing
4	pail	___	rubber bag that
5	veteran		is filled with air
6	ward		

1	blend		
2	devise	___	mix together
3	hug	___	plan or invent
4	lease	___	hold tightly in
5	plague		your arms
6	reject		

1	alcohol	___	stage of
2	apron		development
3	hip	___	state of
4	lure		untidiness or
5	mess		dirtiness
6	phase	___	cloth worn in front to protect your clothes

1	abolish	___	bring to an end by law
2	drip		
3	insert	___	guess about the future
4	predict		
5	soothe	___	calm or comfort
6	thrive		someone

1	apparatus	___	expression of
2	compliment		admiration
3	ledge	___	set of
4	revenue		instruments or
5	scrap		machinery
6	tile	___	money received by the Government

1	bleed		
2	collapse	___	come before
3	precede	___	fall down suddenly
4	reject	___	move with quick
5	skip		steps and jumps
6	tease		

1	bulb	___	female horse
2	document	___	large group of
3	legion		soldiers or
4	mare		people
5	pulse	___	a paper that
6	tub		provides information

1	casual	___	sweet-smelling
2	desolate	___	only one of its
3	fragrant		kind
4	radical	___	good for your
5	unique		health
6	wholesome		

1	concrete	___	circular shape
2	era	___	top of a
3	fibre		mountain
4	loop	___	a long period of
5	plank		time
6	summit		

1	gloomy		
2	gross	___	empty
3	infinite	___	dark or sad
4	limp	___	without end
5	slim		
6	vacant		

10000 WORD LEVEL

1	antics	___	foolish
2	batch		behaviour
3	connoisseur	___	a group of
4	foreboding		things
5	haunch	___	person with a
6	scaffold		good knowledge of art or music

1	acquiesce	___	to accept without
2	bask		protest
3	crease	___	sit or lie enjoying
4	demolish		warmth
5	overhaul	___	make a fold on
6	rape		cloth or paper

1	auspices	___	confused
2	dregs		mixture
3	hostage	___	natural liquid
4	jumble		present in the
5	saliva		mouth
6	truce	___	worst and most useless parts of anything

1	blaspheme	___	slip or slide
2	endorse	___	give care and food
3	nurture		to
4	skid	___	speak badly about
5	squint		God
6	straggle		

1	casualty	___	someone killed
2	flurry		or injured
3	froth	___	being away from
4	revelry		other people
5	rut	___	noisy and happy
6	seclusion		celebration

1	clinch		
2	jot	___	move very fast
3	mutilate	___	injure or damage
4	smoulder	___	burn slowly
5	topple		without flame
6	whiz		

1	apparition		
2	botany	___	ghost
3	expulsion	___	study of plants
4	insolence	___	small pool of
5	leash		water
6	puddle		

1	auxiliary		
2	candid	___	bad tempered
3	luscious	___	full of self-
4	morose		importance
5	pallid	___	helping, adding
6	pompous		support

1	arsenal	___	happiness
2	barracks	___	difficult
3	deacon		situation
4	felicity	___	minister in a
5	predicament		church
6	spore		

1	dubious		
2	impudent	___	rude
3	languid	___	very ancient
4	motley	___	of many different
5	opaque		kinds
6	primeval		

ACADEMIC VOCABULARY

1	benefit		
2	labour	___	work
3	percent	___	part of 100
4	principle	___	general idea
5	source		used to guide
6	survey		one's actions

1	achieve		
2	conceive	___	change
3	grant	___	connect together
4	link	___	finish successfully
5	modify		
6	offset		

1	element	___	money for a
2	fund		special purpose
3	layer	___	skilled way of
4	philosophy		doing something
5	proportion	___	study of the
6	technique		meaning of life

1	convert		
2	design	___	keep out
3	exclude	___	stay alive
4	facilitate	___	change from one
5	indicate		thing into another
6	survive		

1	consent	___	total
2	enforcement	___	agreement or
3	investigation		permission
4	parameter	___	trying to find
5	sum		information
6	trend		about something

1	anticipate	___	control something
2	compile		skilfully
3	convince	___	expect something
4	denote		will happen
5	manipulate	___	produce books and
6	publish		newspapers

1	decade		
2	fee	___	10 years
3	file	___	subject of a
4	incidence		discussion
5	perspective	___	money paid for
6	topic		service

1	equivalent		
2	financial	___	most important
3	forthcoming	___	concerning sight
4	primary	___	concerning money
5	random		
6	visual		

1	colleague	___	action against
2	erosion		the law
3	format	___	wearing away
4	inclination		gradually
5	panel	___	shape or size of
6	violation		something

1	alternative	___	last or most
2	ambiguous		important
3	empirical	___	something
4	ethnic		different that can
5	mutual		be chosen
6	ultimate	___	concerning people from a certain nation

Appendix C: Speaking Test

(Pre-test)

Read the instruction carefully. You are given two (2) minutes to prepare your response. You have two (2) minutes to present your views.

As a student, we get tired easily due to reading, lectures, exams etc. However, they are many ways to overcome tiredness:

- Enough sleep or rest
- Healthy eating habits
- Exercise regularly

1. You are required to explain ways to overcome tiredness based on the points given.
2. In your opinion, which one is the best way to overcome tiredness? Why?

(Post-test)

Read the instruction carefully. You are given two (2) minutes to prepare your response. You have two (2) minutes to present your views.

What helps a person attain happiness in life? Some of the contributing factors are:

- Have a goal
- Good relationship with other
- Healthy lifestyle

1. You are required to explain these factors.
2. In your opinion, which of these factors is the most important factor to help a person attain happiness in life? Why?

Appendix D: Pre- and Post-test Questionnaire

Researcher:
Muhammad Yasir Yahya
Email: m.y.yahya@pgr.reading.ac.uk

QUESTIONNAIRE

You will be asked to complete a short questionnaire about your motivation towards learning and speaking English. The questionnaire should take no more than 10 minutes.

It is entirely up to you whether you participate. You may also withdraw your consent to participate at any time during the project. The information you give will remain STRICTLY CONFIDENTIAL and will only be used for this research project. Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications.

PART A: Demographic Information

Mark in the appropriate space provided.

1. Age: _____

2. Gender

Male	
Female	

3. Ethnicity

Malay	
Chinese	
Indian	
Others: (please state)	

4. What is your MUET Result?

Band 6	
Band 5	
Band 4	
Band 3	

Band 2	
Band 1	
Not applicable	

5. In general, how often do you speak English (either inside or outside the classroom)?

Every day (7 days a week)	
Often (4-5 days a week)	
Sometimes (2-3 days a week)	
Rarely (1-2days a week)	
Never (none)	

6. In your opinion, how would you rate your English speaking performance?

Excellent (81-100%)	
Good (61 - 80%)	
Average (41- 60%)	
Fair (21 - 40%)	
Poor (0 -20%)	
Do not know	

PART B: Motivation towards Learning and Speaking English

This section consists of 45 statements concerning on your motivation and perception towards the learning and speaking English. Indicate the degree to which the statements apply to you under the appropriate number:

1=Strongly Disagree

2=Disagree

3=Agree

4=Strongly Agree

No.	Statements	1	2	3	4
1	I look forward to the time I spend in English class.				
2	Speaking English is important because it will make me more educated.				
3	I do not feel enthusiastic to come to class when English is being taught.				
4	I like to give opinions during English lessons.				

5	Being good at speaking English will help me study other subjects well.				
6	I feel excited when I communicate in English with others.				
7	I like to participate in any classroom speaking activities.				
8	I do not get anxious when I have to answer a question in my English class.				
9	I have more knowledge and more understanding when speaking English.				
10	I am able to focus my attention when speaking English.				
11	Speaking English is very enjoyable.				
12	Speaking English effectively is an important goal in my life.				
13	Speaking in English helps me getting new information in which I can link to my previous knowledge.				
14	Speaking English makes me feel worried.				
15	When someone speaks English well, I like to practice speaking with him/her.				
16	I prefer speaking in my mother tongue rather than English.				
17	Honestly, I only speak English just to pass the speaking test.				
18	I am interested in doing speaking activities in English.				
19	I like speaking English so much; I wish to attend more English classes in the future.				
20	In my opinion, people who speak English are very knowledgeable.				
21	I do not like speaking English because it is difficult for me.				
22	I avoid speaking English when the teacher asks questions.				
23	I feel ashamed if I speak English incorrectly.				
24	I postpone my English homework as much as possible.				
25	I am not physically comfortable whenever I have to speak in my English				
26	Speaking in English helps me communicate and understand better with other people.				
27	I am calm when responding to others in English.				

28	I willingly ask questions to the teacher if I am unclear of what is being taught.				
29	Speaking English makes me have more confidence in expressing myself.				
30	Speaking English helps me to have good relationships with friends.				
31	I like to practice speaking English the way native speakers do.				
32	I cannot apply the knowledge from English subject in my real life.				
33	Speaking English gives me opportunities to meet new people.				
34	I like to befriend people who speak English.				
35	I feel proud when I respond to a question in English correctly.				
36	I am able to think critically when I express my ideas using English				
37	When I miss English class, I never ask my friends or teachers for the homework on what has been taught.				
38	In my opinion, the English language is difficult and complicated to learn.				
39	I have little interest when the teacher is teaching English.				
40	I prefer to be quiet if I do not understand the topic.				
41	Speaking English gives me the opportunity to explore many fields of knowledge.				
42	I feel embarrassed to speak English in front of other students.				
43	I do not pay any attention when my English teacher is explaining the lesson.				
44	I am not satisfied with my English speaking performance.				
45	I can be successful in work and life if I speak good English.				

Appendix E: Pre- and Post-test Interview

Interview Questions

1. Do you find speaking English interesting?
2. Do you think you can pay attention when someone is speaking in English? Why?
3. Would you avoid speaking when someone speaks or ask questions in English? Why?
4. Do you think English helps you to have a good relationship with other people? Why?
5. Do you think speaking English is for the sake of passing exams? Why?
6. Would you attend other English courses in the future to improve your proficiency?
7. How do you feel when speaking in English in front of others? Why?
8. How do you find it when you were speaking in English? Why?
9. Do you think you can speak to others without using Malay? Why?
10. Do you think you can ask questions or express your ideas in English? Why so?
11. In your opinion, do you think someone who speaks English is more knowledgeable?
12. How was English learnt in the classroom? And was speaking skills emphasized in the classroom?
13. How was your teacher like when teaching English? Was the teacher controlling the classroom?
14. Do you think, feedback from the teacher is important? Why?
15. What causes you to participate in the speaking task activities? Why?

Appendix F: Ethical Approval (IoE, University of Reading)

University of Reading

University of Reading
Institute of Education

Ethical Approval Form A
(Version May 2015)

Tick one: Staff project __ PhD ✓ EdD __

Name of applicant (s): Muhammad Yasir Yahya

Title of project: The Development and Implementation of a Speaking Module on Undergraduate Engineering Students in Malaysia

Name of supervisor (for student projects): Dr. Louise Courtney

Please complete the form below including relevant sections overleaf.

Have you prepared an Information Sheet for participants and/or their parents/carers that:	**YES**	**NO**
a) explains the purpose(s) of the project	✓	
b) explains how they have been selected as potential participants	✓	
c) gives a full, fair and clear account of what will be asked of them and how the information that they provide will be used	✓	
d) makes clear that participation in the project is voluntary	✓	
e) explains the arrangements to allow participants to withdraw at any stage if they wish	✓	
f) explains the arrangements to ensure the confidentiality of any material collected during the project, including secure arrangements for its storage, retention and disposal	✓	
g) explains the arrangements for publishing the research results and, if confidentiality might be affected, for obtaining written consent for this	✓	
h) explains the arrangements for providing participants with the research results if they wish to have them	✓	
i) gives the name and designation of the member of staff with responsibility for the project together with contact details, including email . If any of the project investigators are students at the IoE, then this information must be included and their name provided	✓	

k) explains, where applicable, the arrangements for expenses and other payments to be made to the participants	✓		
j) includes a standard statement indicating the process of ethical review at the University undergone by the project, as follows: 'This project has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct'.	✓		
k)includes a standard statement regarding insurance: "The University has the appropriate insurances in place. Full details are available on request".	✓		
Please answer the following questions	**YES**	**NO**	
1) Will you provide participants involved in your research with all the information necessary to ensure that they are fully informed and not in any way deceived or misled as to the purpose(s) and nature of the research? (Please use the subheadings used in the example information sheets on blackboard to ensure this).	✓		
2) Will you seek written or other formal consent from all participants, if they are able to provide it, in addition to (1)?	✓		
3) Is there any risk that participants may experience physical or psychological distress in taking part in your research?		✓	
4) Have you taken the online training modules in data protection and information security (which can be found here: http://www.reading.ac.uk/internal/imps/Staffpages/imps-training.aspx)?	✓		
5) Have you read the Health and Safety booklet (available on Blackboard) and completed a Risk Assessment Form to be included with this ethics application?	✓		
6) Does your research comply with the University's Code of Good Practice in Research?	✓		
	YES	**NO**	**N.A.**
7) If your research is taking place in a school, have you prepared an information sheet and consent form to gain the permission in writing of the head teacher or other relevant supervisory professional?	✓		
8) Has the data collector obtained satisfactory DBS clearance?			✓
9) If your research involves working with children under the age of 16 (or those whose special educational needs mean they are unable to give informed consent), have you prepared an information sheet and consent form for parents/carers to seek permission in writing, or to give parents/carers the opportunity to decline consent?			✓

10) If your research involves processing sensitive personal data[1], or if it involves audio/video recordings, have you obtained the explicit consent of participants/parents?	✓		
11) If you are using a data processor to subcontract any part of your research, have you got a written contract with that contractor which (a) specifies that the contractor is required to act only on your instructions, and (b) provides for appropriate technical and organisational security measures to protect the data?			✓
12a) Does your research involve data collection outside the UK?	✓		
12b) If the answer to question 12a is "yes", does your research comply with the legal and ethical requirements for doing research in that country?	✓		
13a) Does your research involve collecting data in a language other than English?		✓	
13b) If the answer to question 13a is "yes", please confirm that information sheets, consent forms, and research instruments, where appropriate, have been directly translated from the English versions submitted with this application.			
14a. Does the proposed research involve children under the age of 5?		✓	
14b. If the answer to question 14a is "yes": My Head of School (or authorised Head of Department) has given details of the proposed research to the University's insurance officer, and the research will not proceed until I have confirmation that insurance cover is in place.			✓
If you have answered YES to Question 3, please complete Section B below			

Please complete **either** Section A **or** Section B and provide the details required in support of your application. Sign the form (Section C) then submit it with all relevant attachments (e.g. information sheets, consent forms, tests, questionnaires, interview schedules) to the Institute's Ethics Committee for consideration. Any missing information will result in the form being returned to you.

Section A: My research goes beyond the 'accepted custom and practice of teaching' but I consider that this project has **no** significant ethical implications. (Please tick the box.)	✓
Please state the total number of participants that will be involved in the project and give a breakdown of how many there are in each category e.g. teachers, parents, pupils etc. There will be a total of 150 to 300 students involved in the online questionnaire.	

[1] Sensitive personal data consists of information relating to the racial or ethnic origin of a data subject, their political opinions, religious beliefs, trade union membership, sexual life, physical or mental health or condition, or criminal offences or record.

Give a brief description of the aims and the methods (participants, instruments and procedures) of the project in up to 200 words noting:

1. Improving Speaking Proficiency and L2 Motivation through Task-Based Language Teaching (TBLT) on Malaysian Undergraduate Students.
2. The study is being conducted at the university as part of Muhammad Yasir Yahya's PhD thesis. Its aim is to gain a greater understanding on the effects of TBLT on students' speaking performance and and their motivation towards learning and speaking English. A speaking intervention will be developed in improving students' speaking performance and their motivation towards English. It hopes to make recommendations regarding how teachers can teach undergraduate students to ensure that they will be more confident conversing in English.
3. The study is a quasi-experimental study and adopts both qualitative and quantitative methods. In this study instruments such as, questionnaire, interview and speaking test will be used. It contains three stages in which the first stage involve piloting the instruments. The second stage only involves designing and developing the speaking intervention. The last stage will include interview, questionnaire, observation and speaking test. However, the first stage will be conducted initially (December 2015 – February 2016).
4. The participants of the study are undergraduate engineering students in the age of 20 to 25 years of age and will be recruited through simple random sampling. They will be used as the sample of this study as they are undergraduate engineering students who are taking or had taken English courses at their university. Both gender will be involved and a number of 120 participants will be needed as participants for the questionnaire.
5. A student consent form will be used and attached together with the online questionnaire. A consent form will also be given to the Head of Department of the Language Centre.
6. Students are categorised as adult and consent form from their parents is not needed.
7. The estimated date will start within December 2015 to February 2016 for the online questionnaire.

Section B: I consider that this project **may** have ethical implications that should be brought before the Institute's Ethics Committee.	

Please state the total number of participants that will be involved in the project and give a breakdown of how many there are in each category e.g. teachers, parents, pupils etc.

Give a brief description of the aims and the methods (participants, instruments and procedures) of the project in up to 200 words.

1. title of project
2. purpose of project and its academic rationale
3. brief description of methods and measurements
4. participants: recruitment methods, number, age, gender, exclusion/inclusion criteria
5. consent and participant information arrangements, debriefing (attach forms where necessary)
6. a clear and concise statement of the ethical considerations raised by the project and how you intend to deal with then.
7. estimated start date and duration of project.

Section C: Signature of Applicant

Note: A signature is required. Typed names are not acceptable.

I have declared all relevant information regarding my proposed project and confirm that ethical good practice will be followed within the project.

Signed: **Print Name:** Muhammad Yasir Yahya **Date:** 06/11/2015

STATEMENT OF ETHICAL APPROVAL FOR PROPOSALS SUBMITTED TO THE INSTITUTE ETHICS COMMITTEE

This project has been considered using agreed Institute procedures and is now approved.

Signed: **Print Name:** Andy Kempe **Date:** 13/11/15
(IoE Research Ethics Committee representative)*

* A decision to allow a project to proceed is not an expert assessment of its content or of the possible risks involved in the investigation, nor does it detract in any way from the ultimate responsibility which students/investigators must themselves have for these matters. Approval is granted on the basis of the information declared by the applicant.

Appendix G: Ethical Approval (EPU, Malaysia)

UNIT PERANCANG EKONOMI
Jabatan Perdana Menteri
Blok B5 & B6
Pusat Pentadbiran Kerajaan Persekutuan
62502 PUTRAJAYA
MALAYSIA

Tel : *603-8000 8000*
Laman web : www.epu.gov.my

Ruj. Tuan:
Your Ref.:

Ruj. Kami: UPE 40/200/19/3401
Our Ref.: (7)

Tarikh: 11 May 2017
Date:

Muhammad Yasir bin Yahya
43. Lorong 5/1
Taman Pauh Jaya
13700, Perai
Pulau Pinang
Email :

APPLICATION TO CONDUCT RESEARCH IN MALAYSIA

With reference to your application, I am pleased to inform that your application to conduct research in Malaysia has been approved by the **Research Promotion and Co-ordination Committee, Economic Planning Unit, Prime Minister's Department.** The details of the approval are as follows:

Researcher's name	:	**MUHAMMAD YASIR BIN YAHYA**
Passport No./ I.C No	:	**890717-43-5085**
Nationality	:	**MALAYSIAN**
Title of Research	:	**"IMPROVING SPEAKING PERFORMANCE: THE IMPLEMENTATION OF A SPEAKING MODULE FOR UNDERGRADUATE ENGINEERING STUDENTS IN MALAYSIA"**
Period of Research Approved	:	**2 years and 6 months (12.5.2017 – 18.11.2019)**

2. Please take note that the study should avoid sensitive issues pertaining to local values and norms as well as political elements. At all time, please adhere to the conditions stated by the code of conduct for researchers as attached.

3. The issuance of the research pass is also subject to your agreement on the following:

a) to ensure submission of a brief summary of your research findings on completion of your research;

b) to submit three (3) copies of your final dissertation/publication; and

c) to renew your research pass annually.

4. Thank you for your interest in conducting research in Malaysia and wish you all the best in your future research endeavor.

Yours sincerely,

(MUHAMMAD JAWAD BIN TAJUDDIN)
Macroeconomics Section
for Director General
Economic Planning Unit
Prime Minister's Department
Email:
Tel : 03 88723254
Fax : 03 88883798

ATTENTION

This letter is only to inform you the status of your application and **cannot be used as a research pass.**

c.c

Ketua Setiausaha
Kementerian Pendidikan Tinggi
Bahagian Perancangan Penyelidikan dan Penyelarasan Dasar
Aras 13, No. 2, Menara 2
Jalan P5/6, Presint 5
62200 Putrajaya
(u.p Puan Nor Salimah binti Musa)

Appendix H: Information Sheet and Consent Form (Head of Centre)

Researcher:
Muhammad Yasir Yahya
Email: m.y.yahya@pgr.reading.ac.uk

Supervisor
Dr. Louise Courtney
Email: l.m.courtney@reading.ac.uk

HEAD OF CENTRE INFORMATION SHEET

Research Project: *Improving Speaking Proficiency and L2 Motivation through Task-Based Language Teaching on Malaysia Undergraduate Sudents*

Dear Professor,

I am writing to invite your centre to take part in a research study on improving undergraduate students' speaking performance and second language (L2) motivation.

What is the study?

The study will be conducted at the public engineering/technical university as part of Muhammad Yasir Yahya's PhD thesis. Its aim is to gain a greater understanding on undergraduate engineering speaking performance and their motivation towards learning and speaking English. A speaking intervention will be developed and implemented to improve students' speaking performance and motivation and towards learning and speaking English. It hopes to make recommendations in improving spoken English among undergraduate students to ensure that they will be more confident and motivated conversing in English inside and outside the classroom.

Why has this centre been chosen to take part?

Your centre is being invited to take part in this project because the centre provides English courses to undergraduate engineering students that incorporate a number of oracy activities, and because these students will be taking or has taken Malaysian University English Test (MUET) and English Placement Test (EPT).

Does the school have to take part?

It is entirely up to you whether you give permission for the centre to participate. You may also withdraw your consent to participation at any time during the project, without any

repercussions to you, by contacting the project researcher, Muhammad Yasir Yahya, Email: m.y.yahya@pgr.reading.ac.uk

What will happen if the centre takes part?

With your agreement, participation would involve two classes which would be categorised into two experimental groups. Both groups would go through pre/post-test speaking assessments and classroom observations. For the pre- and post-test, a questionnaire will be administered before class starts which takes 5 minutes without interfering the teaching and learning process. Meanwhile, the speaking test would be conducted outside the classroom time, with one student will be called in a separate room to minimise disruption and a few selected students would be interviewed after the test. The interviews will be conducted by the researcher, Muhammad Yasir Yahya, who has previously considerable experience in working with university students.

For the control group, the existing course module (UHL 2400 / Fundamental of English Language) would be used. Meanwhile, a new speaking intervention will be implemented to the experimental group. However, a newly developed speaking intervention will be implemented which still follows the course curriculum specification (UHL 2400 / Fundamental of English Language) but the teaching approach and speaking activities designed would be different from the existing syllabus to meet the purpose of the study. Both groups would be observed for 8-10 weeks between the two tests. All classroom activities including the teaching/learning process and teacher/students interactions will be recorded by the researcher. A timeline of the duration and the process of the study will be attached together with this information sheet.

If you agree to the centre's participation, the researcher will seek further consent from the Head of English Department, the teachers and students.

What are the risks and benefits of taking part?

The information given by participants in the study will remain strictly confidential and will only be seen by the researcher. Neither you, the students nor the centre will be identifiable in any published report resulting from the study. Information about any individuals of the study will not be shared to anyone.

Participants in similar studies have found it interesting to take part. It is anticipated that the findings of the study will be useful for teachers to teach spoken English and for students to improve their speaking performance.

What will happen to the data?

Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you, the teachers, students or the centre to the study will be included in any sort of report that might be published. Participants will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the researcher will have access to the records. The data will be destroyed securely once the findings of the study are written up, after five years. The results of the study will be presented at national and international conferences, and in written reports and articles. The researcher can send you electronic copies of these publications if you wish.

What happens if I change my mind?

You can change your mind at any time without any repercussions. If you change your mind after data collection has ended, we will discard the centre's data.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you can contact my supervisor, Dr. Louise Courtney, Institute of Education, University of Reading. Email: l.m.courtney@reading.ac.uk

Where can I get more information?

If you would like more information, please contact the researcher, Muhammad Yasir Yahya, Email: m.y.yahya@pgr.reading.ac.uk

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form and send it to the email given above.

This application has been reviewed following the procedures of the University of Reading Research Ethics Committee and has been given a favourable ethical opinion for conduct.

Thank you for your time.

Yours sincerely

Muhammad Yasir Yahya
Doctoral Researcher

Institute of Education
University of Reading

Timeline of the Study

Month	Feb 2017		Mar 2017					Apr 2017				May 2017	
Week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Date / **Day**	13 -17	20 -24	27 -3	6 - 10	13 -17	20 -24	27 -31	3 - 7	10 -14	17 -21	24 -28	1 - 5	15 -19
Mon	**Pre-test**							Mid-term Break					**Post-test**
Tue		**Group A**							**Group A**				
Wed													
Thu		**Group B**							**Group B**				
Fri													
Sat													
Sun													

Indicator

Pre-test: questionnaire, speaking test and interview (selected students) for 1 week**

Intervention: speaking intervention for 8 to10 weeks**

Post-test: questionnaire, speaking test and interview (selected students) for 1 week**

***The days shown for experimental groups in the timeline are not the exact day but just as an example. The researcher will need to discuss with the Head of English Language Department and the teachers involved for suitable time/class.*

HEAD OF CENTRE CONSENT FORM

I have read the Information Sheet about the project and received a copy of it.

I understand what the purpose of the project is and what is required of me. All my questions have been answered.

Name of Head of Centre: _____

Name of Centre**:** _____

Please tick as appropriate:

I consent to the involvement of my centre in the project as outlined in the Information Sheet. ☐

Signed: _____

Date: _____

Appendix I: Information Sheet and Consent Form (Students)

Researcher:
Muhammad Yasir Yahya
Email: m.y.yahya@pgr.reading.ac.uk

Supervisor
Dr. Louise Courtney
Email: l.m.courtney@reading.ac.uk

STUDENT INFORMATION SHEET

Research Project: *Improving Speaking Proficiency and L2 Motivation through Task-Based Language Teaching on Malaysia Undergraduate Sudents*

Dear Student,

I am writing to invite your centre to take part in a research study on improving undergraduate students' speaking performance and second language (L2) motivation.

What is the study?

The study will be conducted at the public engineering/technical university as part of Muhammad Yasir Yahya's PhD thesis. Its aim is to gain a greater understanding on undergraduate engineering speaking performance and their motivation towards learning and speaking English. A speaking intervention will be developed and implemented to improve students' speaking performance and motivation and towards learning and speaking English. It hopes to make recommendations in improving spoken English among undergraduate students to ensure that they will be more confident and motivated conversing in English inside and outside the classroom.

Why have I been chosen to take part?

You have been invited to take part in this project because you are a first year undergraduate student who is taking UHL 2400 / Fundamental of English Language.

Do I have to take part?

It is entirely up to you whether you participate. You may also withdraw your consent to participation at any time during the project, without any repercussions to you, by contacting the project researcher, Muhammad Yasir Yahya, Email: m.y.yahya@pgr.reading.ac.uk

What will happen if I take part?

With your agreement, participation would involve answering questionnaire and speaking test, being interviewed and classroom observation. Questionnaire would be administered to you by the researcher before class starts which takes 5 minutes without interfering the teaching and learning process. Meanwhile, the speaking test would be conducted outside the classroom time, with one student will be called in a separate room to minimise disruption and a few selected students would be interview after the test. They would be conducted by the researcher, Muhammad Yasir Yahya, who has previously considerable experience in working with university students. All classroom activities including the teaching/learning process and teacher/students interactions will be recorded by the researcher.

What are the risks and benefits of taking part?

The information given by participants in the study will remain strictly confidential and will only be seen by the researcher. Neither you, your friends, your teachers nor the centre will be identifiable in any published report resulting from the study. Information about any individuals of the study will not be shared to anyone.

Participants in similar studies have found it interesting to take part. It is anticipated that the findings of the study will be useful for teachers to teach spoken English and for students to improve their speaking performance.

What will happen to the data?

Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you, the teachers or the centre to the study will be included in any sort of report that might be published. Participants will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the researcher will have access to the records. The data will be destroyed securely once the findings of the study are written up, after five years. The results of the study will be presented at national and international conferences, and in written reports and articles. The researcher can send you electronic copies of these publications if you wish.

What happens if I change my mind?

You can change your mind at any time without any repercussions. If you change your mind after data collection has ended, we will discard the centre's data.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you can contact my supervisor, Dr. Louise Courtney, Institute of Education, University of Reading. Email: l.m.courtney@reading.ac.uk

Where can I get more information?

If you would like more information, please contact the researcher,

Muhammad Yasir Yahya, Email: m.y.yahya@pgr.reading.ac.uk

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form and send it to the email given above.

This application has been reviewed following the procedures of the University of Reading Research Ethics Committee and has been given a favourable ethical opinion for conduct.

Thank you for your time.

Yours sincerely,

Muhammad Yasir Yahya
Doctoral Researcher

Institute of Education
University of Reading

STUDENT CONSENT FORM

I have read the information above, understand the purpose of the study and what is required of me**.** Please tick as appropriate:

I consent to completing the questionnaire:	**YES**	
	NO	

Signature **:** _____

Name **:** _____

Student ID **:** _____

Date **:** _____