不同類型的線上診斷性文法測驗的解答回饋

對於大一學生文法能力的影響

The Effectiveness of Different Types of Online Diagnostic Grammar Test Feedback on University EFL Freshmen's Grammar Proficiency

by

詹雅婷 Ya-Ting Lily Chan

THESIS

Presented to the Faculty of the

Department of Foreign Languages and Literature of

Tunghai University

in Partial Fulfillment of the Requirements for the Degree of

MASTER OF ARTS in

Teaching English As A Foreign Language

TUNGHAI UNIVERSITY

May 2012

中 華 民 國　一百零一　年　五　月

Acknowledgements

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed their valuable assistance in the preparation and completion of this study.

First and foremost, I offer my utmost gratitude to my thesis adviser, Dr. James Sims, who has helped and guided me not only with the writing tasks but also with the data collection and analysis procedures. He was always generous in contributing his expertise and ideas to help me solve problems in the research process. He was always so patient, kind, and humorous. I will never forget him.

I would also like to express my sincere gratitude to my thesis co-adviser, Dr. Jyu-Fang Annie Yu, who I first met in my junior year in undergraduate program and who later became a very important person in my life. Dr. Yu not only has supervised my thesis, but also has been caring for me for the past five years. She is the person who most cares about my studies and my life. She has offered me innumerable suggestions, support, encouragement, and help. My gratitude to her can never be expressed in mere words.

I would also like to thank my committee member, Dr. Chih-Hui Chang, for her contribution in giving suggestions and comments on my thesis. She was always very kind

and patient each time we contacted each other by e-mail and at the proposal meeting. Her expertise and generous help made my thesis become better.

In addition, my gratitude goes to Dr. Muchun Yin, my former adviser, who has given me inspiration on the thesis, suggestions and guidance on my writing of literature review. His assistance has helped me greatly at the initial stage of this study.

Special thanks to teachers from our MA program and FENM teachers from the English Center. Without their assistance, this research could not have been conducted successfully. I would especially like to express my gratitude to FENM teacher Candace Yu, who was a great help to me with the pilot study and the data collection procedures.

I would also like to thank the participants of this study, for their cooperation, patience, and concentration in the data collection process.

My classmates and friends in the MA program also deserve a mention. Ginna Chi, who has been a good friend since we were undergraduates, supported me from the beginning to the end of my thesis, and also assisted me in the data collection procedures.

I would like to thank the Chung Hwa Rotary Educational Foundation for their scholarship that gave me financial support.

Last but not least, my family, and God. I thank my mother for her consideration, encouragement, and financial support. She never gave me pressure but always supported

me in my graduate school years. She never pushed me to study harder, but only hoped that I could live a happy and fulfilling life. I thank God for answering my prayers and giving me the strength, courage and confidence to overcome difficulties in life.

Ya-Ting Lily Chan

不同類型的線上診斷性文法測驗的解答回饋

對於大一學生文法能力的影響

研究生: 詹雅婷

指導教授: 幸雅各 博士
尤菊芳 博士

摘要

過去有些學者研究不同類型之線上診斷性文法測驗解答回饋對外語文法學習的功用，這些研究主要使用的語言不全然是英文，也未必是在探索亞洲地區學生的表現。因此，以台灣地區為例，目前並沒有找到實驗研究報告，來支持哪一類型的解答回饋最為有效，以讓教師以及學生應用於文法學習上。

本項研究旨在探討不同類型之線上診斷性文法測驗的解答回饋，對於非外文系大一新生的文法能力進步是否有所影響。不同類型線上診斷性文法測驗的解答回饋有效度的探究，能夠使教師了解哪一類型的解答最為有效，進而用來幫助學生達到較為有效的學習。

本項研究採用前測/後測設計。共有九十位大一新生來自三個不同程度(高、中低）的大一英文班級，參與此項研究。參與測驗者被平均分成一個實驗組，以及一個對照組。在數據分析方面，兩個獨立樣本 T 檢定及兩個成對樣本 T 檢定分別被用來分析在實驗組及對照組之間及內部前後測平均數比較。

研究分析結果顯示，有接受詳解(metalinguistic feedback) 的實驗組學生，在後測的表現上進步許多。相較之下，對照組學生(沒有接受詳解)並無顯著的進步。

***關鍵字*:** 文法能力、診斷性文法測驗、後設語言解答回饋、前測、後測

The Effectiveness of Different Types of Online Diagnostic Grammar Test Feedback on University EFL Freshmen's Grammar Proficiency

Name: Ya-Ting Lily Chan

Thesis Advisers: Dr. James Sims
Dr. Jyu-Fang Yu

ABSTRACT

Several research on the effectiveness of different types of feedback in diagnostic grammar tests have been conducted previously (Nagata, 1993; Heift, 2004; Bitchener & Knoch, 2008; Bitchener *et al*, 2005; Bitchener, 2008; Ellis *et al*, 2006, E.E. Jang, 2009). However, little research has been done in Asia, and the target languages in some of those studies were not even English. There are not enough findings on the effectiveness of different types of feedback in grammar testing for Taiwanese teachers and students to act upon.

The current study aims to investigate whether the type of feedback (no corrective feedback; metalinguistic feedback) on an online diagnostic multiple-choice grammar test results in improvement in EFL university freshmen's grammar proficiency. Investigation of the effectiveness of different types of online diagnostic grammar feedback on students' grammar proficiency helps instructors understand what types of grammar feedback are

the most useful and further apply it in class to help students achieve facilitative learning.

The present study adopted the pre-/ post- test design. Three different levels of FENM classes (one high, one mid, and one low) students were involved in this study (N=90). The three FENM classes were divided into one control group and one treatment group. Two independent-sample t-tests were used to analyze the mean scores of pre-tests and post-tests between groups, while two paired-sample t-tests were used to analyze the mean scores of pre-tests and post-tests within groups.

Data analysis revealed that control group students did not improve significantly on the post-test, while treatment group students improved greatly on the post-test. The results indicated that treatment (metalinguistic feedback) was the main factor that could explain the differences between the groups in the current study.

Key words: grammar proficiency; diagnostic grammar test; metalinguistic feedback; pre-/post-test

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Introduction

In EFL (English as a Foreign Language) settings, measuring language proficiency is crucial because it is regarded as the criteria for assessing learners' language skills (Alderson, 2005). Researchers have different preferences for the use of *language proficiency* and *language ability*. For example, Brown (2004) preferred the term *proficiency*, while Bachman & Palmer (1996) preferred the term *ability*. Though there is no consensus on whether to use *proficiency* or *ability* to describe what language learners can do, both of these terms consist of various related constructs that can be specified and measured. For example, *language ability/proficiency* consists of separate components embodied in the four skills: listening, speaking, reading and writing (Bachman & Palmer, 1996).

Grammar learning is regarded as essential in foreign language acquisition. Purpura (2004) points out that to know a language means to be able to apply the rules of grammar (implicitly or explicitly). On the interfaces of grammar leaning, Dekeyser (1998; 2003) argued for the strong position in which explicit grammatical knowledge can be converted into implicit learning directly, while Doughty and Williams (1998) and Ellis (2002) argued for the weak position under which explicit grammatical knowledge cannot be

converted into implicit learning directly, but only facilitates learning development. To understand the relationship between explicit grammar knowledge and implicit learning, researchers have investigated whether noticing and consciousness-raising promote better learning (e.g. Leow, 1997, 2000; Mackey, 2006; Izumi, 2002; Sheen, 1992; Fotos, 1993; Sugiharto, 2006). The current study takes Dekeyser's (1998; 2003) position of the strong interface in which explicit grammatical knowledge can convert into implicit learning directly.

For teachers and students, knowing how much students have learned or have not learned is important. Both learning and teaching could be adjusted based on information elicited from the test results (Black & Wiliam, 1998; Alderson, 2005; Yin, in press). In order to facilitate learning efficiency through testing, *diagnostic language testing* is often designed to provide not only feedback which identifies the strengths and weaknesses of learners' knowledge of the target language but also information on what they need to work on in the future (Brown, 2010; Alderson, 2005). Furthermore, with the advent of computer-based tests, online diagnostic language testing can be applied in many educational settings, and the feedback can be provided immediately to maximize the impact on the test-takers' inter-language development (Alderson, 2005; Chapelle & Douglas, 2006).

It is argued that computer-assisted diagnostic language testing facilitates learning efficiency by providing immediate feedback (Black & Wiliam, 1998; Brown, 2004; Alderson, 2005; Chapelle & Douglas, 2006). Phases of online diagnostic grammar testing and types of feedback are worth exploring. Researchers have examined the effectiveness of types of computer feedback in online grammar tests (e.g. Nagata, 1993; Heift, 2004; Bitchener & Knoch, 2008; Bitchener *et al*, 2005; Bitchener, 2008; Ellis *et al*, 2006, E.E. Jang, 2009). Further, investigation of the effectiveness of different types of online diagnostic feedback on students' grammar proficiency helps teachers understand students' strengths and weaknesses, offers the teachers implications for their pedagogy, and enables testing to become an instrument for learning.

However, few studies have been conducted on the effectiveness of different types of diagnostic grammar feedback. Furthermore, in previous research conducted on the effectiveness of diagnostic grammar test feedback, the usefulness of feedback was usually based on test-takers' subjective perceptions and thus lacked independent evidence of its effects (Yin *et al*, in press). Therefore, the current study attempts to identify the effectiveness of different types of online diagnostic grammar test feedback by comparing students' scores to investigate their improvement between the pre-test and the post-test.

1.1 Background of the Study

The education system in Taiwan is more exam-oriented than in western educational settings and students need to take entrance exams at different stages of school life. Therefore, how to help students learn from taking tests becomes crucial in Taiwan's educational setting.

In order to prepare for paper-and-pencil English exams at different stages of school life, Taiwan's English education still focuses on traditional grammar teaching (Yeh, 2004; Tsai, 1998). In the past, English assessments were usually designed for placement purposes. Benefits of assessment such as facilitation of learning efficacy were ignored (Alderson, 2005; Yin *et al*, in press). In fact, in addition to its placement purpose, assessments could be used as a way to better understand learners' strengths and weaknesses (Brown, 2010; Alderson, 2005). Further, Purpura (2004) points out that feedback helps learners to connect the target language structure to their inter-language by making learners *notice*. The researcher of the current study would like to investigate whether diagnostic grammar testing with feedback in classrooms is able to promote facilitative learning. Therefore, the effectiveness of types of feedback on FENM students' grammar proficiency is a topic worthy of research.

Moreover, with the development of technology, computer-based tests have been integrated into classrooms to achieve more efficient learning than traditional approaches.

According to Canale (1986), use of computer technology provides better means for measuring different language constructs than traditional methods (Chapelle & Douglas, 2006). Moreover, Alderson (2005) points out that the computer can present information in a variety of ways and can encourage the learner's own strategies for evaluation (Chapelle & Douglas, 2006). Hence, computer technology not only provides learners with an informative platform but also allows the test-takers to self-assess their knowledge and facilitate their inter-language development (Alderson, 2005; Black & Wiliam, 1998, Purpura, 2004).

1.2 Statement of the Problem

Issues and phases of assessing foreign language proficiency and grammar learning in EFL settings have been discussed by scholars (e.g. Black& Wiliam,1998 ; Purpura, 2004; Alderson, 2005; Brown, 2010). In the case of interfaces of grammar knowledge (see Chapter 2), empirical studies have explored the relationship between explicit grammatical knowledge and implicit learning (Leow, 1997, 2000; Mackey, 2006; Izumi, 2002; Sheen, 1992; Fotos, 1993; Sugiharto, 2006). Researchers have also examined the effectiveness of different types of feedback in diagnostic grammar tests (Nagata, 1993; Heift, 2004; Bitchener & Knoch, 2008; Bitchener *et al*, 2005; Bitchener, 2008; Ellis *et al*, 2006, E.E. Jang, 2009). However, little research has been done in Asia, and the target languages in

some of those studies were not English. There are not enough robust findings of the effectiveness of types of feedback in grammar testing for Taiwanese teachers and students to act upon.

Furthermore, since exams play a powerful role in education in Taiwan (Chen *et al*, 2005), research on the usefulness of feedback on diagnostic grammar testing is important because if students could learn from taking tests, learning is assumed to be efficient.

1.3 Purpose of the Study and Research Questions

The current study aims to investigate the effects of an online diagnostic multiple-choice grammar test and the provision of different levels of online grammar feedback on learners' grammar proficiency. The researcher seeks to learn how strongly the level of detail in the feedback, such as providing metalinguistic explanations, is related to the development of learners' grammar knowledge. Therefore, the present study addresses the following research questions.

1. Are there any statistical differences in EFL university freshmen's grammar performance in terms of their mean scores on the online grammar tests within the control group and within the treatment group which received metalinguistic feedback?

2. Are there any statistical differences in EFL university freshmen's grammar performance in terms of their mean sores on the post-test between the control group and the treatment group?

1.4 Definition of Terms

In order to ensure clear understanding of the frequently used terms through the literature review and to avoid any confusion, the following definitions are provided.

1. Metalinguistic awareness

The term metalinguistic awareness was first used and defined by Cazden (1974) as the transfer of linguistic knowledge and skills across languages. It is the ability to objectify language as a process as well as a thing. For example, people who are able to describe a sentence using linguistic terms such as subject, verb and object have metalinguistic awareness.

2. Implicit/Explicit learning

According to Schmidt (2001), implicit learning refers to *learning without metalinguistic awareness*, while explicit learning refers to *learning with metalinguistic awareness*. Ellis (2005) also points out that implicit learning happens when the learner have internalized a linguistic feature without awareness. By contrast, explicit learning takes place with the learner's awareness of the linguistic feature.

3. *Metalinguistic feedback*

Metalinguistic feedback refers to comments and/or information that point out errors with linguistic terminology without providing the correct form (Purpura, 2004; Ellis, 2008).

1.5 Significance of the Study

This current study attempts to identify the effectiveness of diagnostic grammar test feedback by investigating the improvement after the students receive different types of feedback. The researcher hopes the findings of the current study can help instructors and scholars understand what types of grammar feedback are the most useful and further apply it in class to help students achieve better learning. Moreover, the findings are expected to aid instructors in giving feedback to different levels of students. Furthermore, since grammar learning is essential in learning a language and students are required to take exams frequently in Taiwan's educational system, the researcher hopes to help Taiwanese students learn from taking exams. Finally, the researcher hopes that the computer-based diagnostic tests can be applied in class and further help students learn English grammar efficiently.

Chapter Two

Literature Review

This literature review includes five main sections. The first section (2.1) focuses on grammar learning. In 2.2, the focus shifts to a discussion of assessment of grammar knowledge and learning. Theories and research in diagnostic language testing are then reviewed in 2.3. Types and phases of feedback are covered in section 2.4. Finally, section 2.5 identifies the research gap that the current study addresses.

2.1 Grammar Learning

2.1.1 Types of Grammatical Knowledge and the Strong & Weak Interface Positions

Grammatical knowledge of a language has always been considered to be the core of foreign language proficiency development (Alderson, 2005; Purpura, 2004). Traditionally, having grammatical ability meant being able to recite or recall the construction of grammatically accurate forms and sentences (Alderson, 2005). More recently, stress on grammatical construction of language forms has shifted to grammatically appropriate forms, that is, the meaning of the language. Related to this shift, theorists hold different positions on the interfaces of grammatical knowledge. In this section, two types of

grammatical knowledge, implicit and explicit knowledge, will be described, followed by a discussion of how they interface.

Language learning scholars have made several important claims about the definitions and features of the types of grammatical knowledge. Explicit grammatical knowledge is conscious, declarative, verbalizable, and available for use through controlled processing (Ellis, 2005). Further, Ellis *et al* (2006) point out that explicit grammar knowledge may be linked to metalinguistic labels. By contrast, implicit knowledge is tacit, intuitive, and available for automatic processing (Ellis, 2005). In other words, explicit grammatical knowledge refers to grammar rules that can be articulated, codified and stored; while implicit grammatical knowledge refers to grammar rules that have been internalized by the learner without the learner being aware of it.

In addition, Schmidt (2001) defines implicit learning as learning without metalinguistic awareness, while explicit learning refers to learning with metalinguistic awareness. Ellis (2005) observes that the terms implicit/explicit learning are usually only considered in relation to the learner's perspective. He further states that implicit learning occurs when the learner has internalized a linguistic feature without being aware that he has done so. By contrast, explicit learning happens with the learner's awareness that he has learned the linguistic feature.

In the acquisition of L1 grammar, Ellis (2006) argues that learning takes place from

the experience of usage instead of from the learning of explicit rules. This implicit learning is extracted from natural meaningful communication. In this case, the learner's self-awareness allows reflective examination, re-organization and optimally results in explicit representations in daily communication with others. He goes on to say that when learners acquire the grammar of a second language, exposure to explicit knowledge takes place in advance of implicit knowledge, and it takes effort for explicit knowledge to be converted into implicit knowledge. He further states that, implicit learning is the acquisition of knowledge of the target structure which takes place naturally and without conscious operations. In other words, in implicit learning, the target structure (explicit knowledge) is converted into implicit knowledge subconsciously, without the learner being conscious of the transformation of explicit knowledge into implicit knowledge (Ellis, 2006).

Explicit learning, by contrast, occurs when learners attend to a stimulus environment and requires a more conscious operation (Ellis, 2006). That is, implicit learning requires no learner awareness of the outside stimuli because the intake is achieved naturally. Explicit learning, however, requires learner's awareness of the new concepts from the outside, and learners need to convert these new concepts into intake using learning processes that they are more conscious of. Ellis (2006) also argues that the vast majority of the learner's cognitive processing is unconscious. Taken together, Schmidt (2001),

Ellis (2005) and Ellis (2006) provide arguments for a strong correlation between consciousness and implicit/explicit learning.

In addition to the claims of implicit/explicit knowledge and learning made by scholars, different interfaces of the relationship between explicit knowledge and implicit learning have been posited, and empirical studies have been conducted by several researchers (e.g. Ellis, 2005; Ellis, 2006). Further, since learning and teaching are always tied together across many aspects of language acquisition, whether explicit instruction leads to implicit learning has been debated intensely, and different scholars hold different positions on the issue (see Dekeyser, 1998, 2003; Ellis, 2005).

DeKeyser (1998, 2003) has argued for a strong interface, meaning that explicit knowledge converts into implicit knowledge through communicative practice. He has suggested that "explicit knowledge can be fully automatized and thereby become functionally equivalent to implicit knowledge" (Ellis, 2005, p. 149). In other words, explicit input is able to become intake, which leads to automatic output by learners.

However, the weak interface position holds that explicit knowledge does not convert into implicit knowledge directly; instead, explicit knowledge merely indirectly facilitates learning development by inducing attention to form (Doughty & Williams, 1998; Ellis, 2002; Ellis, 2005).

The current study takes the strong interface position that explicit knowledge can be

converted into implicit learning in grammar acquisition. One reason is that it is currently more accepted by researchers. In addition, there is empirical support for this view. After evaluating the results of research on the effects of explicit instruction, Ellis (2005) concludes that explicit instruction can result in implicit learning because of the "noticing" of language. Moreover, Norris and Ortega (2000) reported a meta-analysis of 49 empirical studies, which concluded that explicit types of instruction were more effective than implicit types and that the effectiveness of L2 instruction was durable (Ellis, 2006). The role of noticing hypothesis and awareness in the learning process will be discussed next.

2.1.2 The Noticing Hypothesis

"What learners notice in input is what becomes intake for learning" (Schmidt 1995, p.20). The Noticing Hypothesis (NH) was proposed by Richard Schmidt in 1990, who defined the term "noticing" as "elements of the surface structure of utterances in the input, instances of language, rather than any abstract rules or principles of which such instances may be exemplars" (Carroll, 2006, p. 17). Schmidt (1990) claims that L2 learners need to become aware of the input before they begin to acquire a language feature. Here, it does not mean that "noticing" would result in acquisition, but "noticing" is the essential starting point, and is vital for L2 inter-language development. "Intake is that part of the input that the learner notices" (Carroll, 2006, p. 17). In other words, learners must first

have conscious awareness of some particular language form in the input, or "notice" it, before they can actually take in the new concepts.

In addition, awareness at the level of noticing is required for learning, while awareness at the level of understanding is facilitative, which leads to deeper and more rapid learning. "SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be" (Schmidt, 2001, Carroll, 2006, p. 17).

Explicit knowledge plays a role in the perception of L2 form by facilitating the process of learners' "noticing" stimuli from the outside and the gap between building their inner knowledge and new concepts (Ellis, 2005; Ellis, 2006).

Several recent empirical studies support the Noticing Hypothesis. Leow (1997, 2000) examined the relationship between noticing, awareness and learning outcomes of adult learners of Spanish in written production tasks targeted at stem changing verbs. These two studies found that those who demonstrated a higher level of awareness, or "understanding", learned the most. However; those who noticed instances but failed to attempt generalization of the changing verbs did not learn as much as those who demonstrated a higher level of awareness.

Additionally, there is evidence showing that there is no learning in the absence of noticing. For instance, Mackey (2006) investigated whether feedback promoted the

noticing of targeted L2 grammatical features (plurals and past tense forms) and whether there was a relationship between learners' reports of noticing and learning outcomes. The findings revealed that learners report more noticing when feedback is provided and learners who noticed more developed more. Further, Izumi (2002) conducted an experimental study to compare the effects of output and enhanced input on noticing and development. This study found that participants who demonstrated more noticing had better learning outcomes.

In sum, "noticing" is regarded as an essential element in second language acquisition (Schmidt, 2001; Ellis, 2005), and is essential in grammar learning (Purpura, 2004; Mackey, 2006). Further, the findings of Leow (1997, 2000), Mackey (2006) and Izumi (2002) all offer evidence to support the Noticing Hypothesis. In sum, the higher the level of noticing, the better the learning performance of learners.

2.1.2.1 Consciousness-Raising

Consciousness-Raising was first introduced by Sharwood-Smith in 1981. It was subsequently discussed in detail by Rutherford (1987), who defines Consciousness-Raising (CR) as "the drawing of the learner's attention to features of the target language" (p. 189). He further points out that when drawing students' attention, Consciousness-Raising (CR) is used as a way to guide the grammar learning process to

make the target features become more familiar to the students. Similarly, C-R was defined by Ellis (2005) as "a deliberate attempt on the part of the teacher to make the learners aware of specific features of the L2; it entails an attempt to instill an understanding of the formal and functional properties of these features by helping the learners develop a cognitive representation of them" (Bhattarai, 2000, p. 15). Through Consciousness-Raising, the development of explicit knowledge of targeted features is raised, the learner's cognitive domain is activated and learning is facilitated, finally leading to implicit learning (Rutherford, 1987).

In addition, as mentioned in 2.1.2, awareness at the level of noticing is necessary for learning, and awareness at the level of understanding will foster deeper and more rapid learning (Schmidt, 2001; Ellis, 2008). Therefore, consciousness plays a crucial role in learning and Consciousness-Raising promotes better grammar learning.

In support of these claims, a number of studies have been conducted to investigate the efficacy of grammar Consciousness-Raising. Sheen (1992) examined the effectiveness of direct and indirect Consciousness-Raising in a six-week beginner's French course for Japanese and found that students in the two groups performed equally well in a written post-test of the structure taught previously. Fotos (1993) investigated the amount of learner noticing produced by two types of grammar Consciousness-Raising treatments, teacher-fronted grammar lessons and interactive, grammar problem-solving tasks. Her

study involved 160 Japanese college students of English divided into two treatment groups, which were taught indirect object placement, adverb placement, and relative clause usage in communicative input. She found that both types of grammar Consciousness-Raising are effective in promoting a significant level of noticing of the target language structures in subsequent communicative input. Finally, Sugiharto (2006) investigated Indonesian students' ability to understand simple present tense rules using a grammatical judgment test. Sugiharto (2006) compared the results from students' pre/post-tests, and found that students performed significantly better on the post-test. This study confirmed that Consciousness-Raising is effective in helping students develop their explicit knowledge of the simple present tense. It appears that Consciousness-Raising effectively helps develop learners' explicit knowledge, promotes the level of noticing, and ultimately facilitates learning.

2.2 Assessing Grammar Knowledge and Learning

2.2.1 Types of Grammar Testing

As mentioned in 2.1, knowledge of grammar is regarded as central to the development of foreign language proficiency (Alderson, 2005). Further, to know a language means to know the language structure (Purpura, 2004). Recently, emphasis on the importance of meaning has increased as communicative approaches have begun to

thrive (Alderson, 2005). As a result, when assessing grammar knowledge, learners are often tested on their ability to express the meaning in the context in grammatically appropriate ways and the assessment of grammatical knowledge is based on tasks requiring students to demonstrate their ability to communicate in speaking or writing (Purpura 2004; Alderson, 2005). According to Purpura (2004), "grammar is seen as a set of rules to be internalized and used for communication" (p.3), and he further states that the tasks test-makers include on tests should match the types of language-use tasks found in real-life or language instructional domains.

Multiple-choice question type and open-ended questions type are used most often in grammar testing. These approaches will be discussed in this section. Additionally, since computer technology has been integrated into language testing in many educational settings, the value of computer-assisted grammar testing will be explored in 2.2.2.

2.2.1.1 Open-Ended Question Type

Compared to multiple-choice tests, the open-ended question approach is not common in grammar testing due to its construction and scoring method. Unlike answers to multiple-choice questions, responses to open-ended questions are limited-production or extended-production tasks (Purpura, 2004; Brown, 2010; Abeywickrama, 2007). They have a degree of subjectivity in scoring because judgment is required to interpret and

evaluate performance against the criteria for correctness. Further, open-ended questions require the test-taker to produce responses that range in length from a word to a sentence or two, or even an essay.

Tasks for open-ended questions include gap-filling activities, short-answer activities, discourse completion activities, summaries and essays, which require the test-taker to provide full information that involves their grammatical competence. Answers that test-takers give in this type of question contain large amounts of information useful in judging to what extent the test-taker has mastered the grammatical rules (Purpura, 2004). However, in testing grammatical rules by means of computer, these type of tasks are not practical because they requires subjective judgment, which cannot be easily integrated into online assessment tools (see Table 2.1).

2.2.1.2 The Multiple-Choice (MC) Type

When learners' explicit knowledge of certain grammatical forms are assessed, multiple-choice tests are often used. They are valid and reliable ways to evaluate learners' grammatical competence (Kitao & Kitao, 1996, Chen *et al*, 2006; Purpura, 2004). Even though multiple choice questions do not necessarily show how much grammatical knowledge the test takers have internalized, nor whether they are able to use them automatically and spontaneously, the results can still show how many of the targeted

grammatical forms students have mastered (Purpura, 2004).

Multiple-choice (MC) questions may be thought of as objective test tasks because they require no expert judgment to evaluate performance with regard to the criteria for correctness (Purpura, 2004). The responses that test-takers give are selected from provided items, meaning that the responses are a form of selected-response task. MC items present input with gaps or underlined words or phrases. The test-takers need to choose the correct, best or most appropriate, acceptable answer from the items given. Only one option contains the correct answer; the incorrect options are known as distractors. In the sample below, D is the correct answer, while A, B, and C are distracters.

She asked me _____
(A) where I to buy my shoes.
(B) where did I buy my shoes.
(C) where bought I my shoes.
(D) where I bought my shoes.

MC items are well suited to testing grammatical forms, especially in computer-assisted language testing because they can be checked objectively (Purpura, 2004). Further, multiple-choice items are fast, easy, and economical to grade. Because they are machine scorable (Bailey, 1998). Their surface objectivity gives the test the appearance of being fairer or more reliable compared to subjectively scored tests such as open-ended question tests (see Table 2.1). Therefore, MC grammar testing is adopted in

this current study.

Table 2.1
Types of Grammar Testing

The Multiple-Choice (MC) Type	Open-Ended Question Type
1) Requires no expert judgment to evaluate performance with regard to the criteria for correctness.	1) Requires expert judgment to interpret and evaluate performance with regard to the criteria for correctness.
2) Present input with gaps or underlined words or phrases. The test-taker need to choose the correct, best or most appropriate, acceptable answer from the items given.	2) Require the test-taker to produce responses that range in length from a word to a sentence or two, or even an essay.
3) Is well suited in testing grammatical form especially in computer-assisted language testing because it can be scored objectively.	3) Requires the test-taker to provide full information elicited from their grammatical competence.
4) Is valid and reliable to evaluate learners' grammatical competence.	4) Is not suitable in testing grammatical rules by means of computer because it requires subjective judgment
5) Is fast, easy, and economical to score.	5) Takes time to score.

2.2.2 The Value of Computer-Assisted Grammar Testing

In a meta-analysis of research on the use of multimedia to teach a variety of subjects, Rogan *et al* (1993) found that, in general, multimedia instruction reduces learning time by 30% compared to traditional instruction (Nutta, 1998). Computer-assisted applications may be particularly effective at facilitating the development of grammatical competence

because such applications are rich sources of comprehensible input (Collentine, 1997a; Collentine, 1998). The use of computer technology provides a better means for measuring language compared to traditional methods of assessment (Canale, 1986, Chapelle and Douglas, 2006). Computers have the ability to measure time, record information about the testee's routes through the test, and encourage the learner's own strategies for evaluation (Alderson, 2005, Chapelle and Douglas, 2006).

Furthermore, as Alderson (2005) points out, computer-assisted language testing can be designed to give learners immediate feedback on their performance, which has maximum impact because immediate feedback helps the learner recall their responses to questions so as to develop their inter-language. He further states that if test construction is based on language use, development, and learning theory, and if learners can receive feedback immediately to help them reflect on their inter-language, the integration of assessment and language learning is possible. In addition to the possibility of offering immediate feedback, the possibility for response analysis and recording for further analysis exceed what is feasible with traditional assessment (Chapelle & Douglas, 2006).

Further, feedback has the potential to become beneficial washback (Brown, 2004). "CALT may hold the potential for a type of positive washback if one considers the benefits that many beliefs are associated with regular access to technology during second language learning"(Chapelle & Douglas, 2006, p.18). Here, another value of

computer-assisted language testing is identified by the researchers mentioned above. Computer-assisted grammar testing has clearly become an integral part of language education settings, and is thus one crucial element of this current study.

2.3 Diagnostic Language Testing

2.3.1 Introduction

Traditional testing such as achievement and proficiency tests are used only to judge learner ability at the current stage. However, one of the main elements of testing, helping students to learn, has been long neglected (Alderson, 2005). Further, traditional testing has been criticized for not providing diagnostic information to inform students of their strengths and weaknesses in a specific academic domain (Nichol, 1994; Snow & Lohman, 1989; Jang, 2009).

According to Brown (2010) and Abeywickrama (2007), achievement tests and proficiency tests only provide summative information instead of formative information that diagnostic tests can provide. Achievement tests are primarily summative because they are administered at the end of a given period of instruction. Further, achievement tests aim to examine whether course objectives have been met and whether appropriate knowledge and skills have been acquired. Proficiency tests, by contrast, test global competence in a language. Proficiency tests are not limited to any one course, curriculum,

or single skill in the language; rather, they test overall ability (Brown, 2004). Achievement tests and proficiency tests only inform test-takers of their knowledge of the target language or test results. However, information on what aspects of their language abilities they need to improve is absent.

Assessing foreign language proficiency is crucial in many language settings. Though instructors know that understanding learners' current knowledge must take place prior to helping them learn, valid types of diagnostic assessment are lacking (Alderson, 2005; Yin, Sims, & Cothran in press). Furthermore, diagnostic language testing helps in eliciting information on what aspects of their languages abilities students need to focus on in the future (Brown, 2004), but few diagnostic tests exist, because they are difficult to construct (Alderson, 2005). Davis *et al* (2003) also point out that "It is difficult and time-consuming to construct a test which provides detailed diagnostic information" (Alderson, 2005, p. 6). However, Hughes (2003) suggests that computer-based testing might offer a solution (Alderson, 2005).

Diagnostic language testing is designed to diagnose specified aspects of a language, and will elicit information on what students need to work on in the future and further offer detailed information on a learner (Brown, 2010). In addition, the ALTE multilingual glossary defines a diagnostic test as "A test which is used for the purpose of discovering a learner's specific strengths or weakness. The results may be used in

making decisions on future training, learning or teaching" (Alderson, 2005, p.4). Schonell & Schonell (1960) argue that errors provide the most information, and give detailed examples (Alderson, 2005).

Therefore, diagnostic tests identify strengths and weaknesses in learners' knowledge and use of language, and are more likely to focus on weaknesses than on strengths (Alderson, 2005). Alderson (2005) goes on to say that so-called weaknesses could be interpreted as the normal inconsistencies between learner's competence and the explicit knowledge of the target language in foreign language development. Heift (2004) also points out that feedback can provide an explanation of the error and highlight the error in the student input.

Hence, if the mismatch between learners' current knowledge and knowledge of the target language could be diagnosed, there is a great possibility for both learners and teachers to plan future language learning and teaching activities to improve the efficacy of foreign language acquisition.

Feedback is the crucial component of any diagnostic test and it provides meaningful information to users to understand. Further, users or their teachers can act upon the feedback in the future (Alderson, 2005). Diagnostic language testing's provision of meaningful feedback was defined by Heift (2003) as "response that provides a learning opportunity for students" (Cotos and Pendar, 2008, p. 533). Further, Cotos and Pendar

(2008) claim that diagnostic tests might enhance learning opportunity by allowing learners to act upon the received feedback, thus make diagnostic assessment interactive.

In addition, with the advent of computer-based tests, the provision of immediate feedback to test takers becomes simple, making it potentially highly informative and relevant (Alderson, 2005; Brown, 2004). In other words, with the integration of computer technology, administration of diagnostic language testing becomes feasible in many educational settings. Furthermore, test results and detailed feedback could be given immediately via well-designed computer programs.

Alderson (2005) describes the DIALANG project as an online diagnostic language testing system which was a European Union-founded project started in 1996 and came to the end of its public founding in 2004. The DIALANG project was set up to diagnose five language skills (Reading, Listening, Writing, Vocabulary and Grammar). This project provides feedback to users rather than certifying their proficiency. The tests are delivered over the Internet and the test-taker's responses are sent back to the servers during the actual session, after which they then receive test results and feedback. The tests are free of charge to users and diagnostic tests are being developed in 14 European languages. Further, the test specifications are based on the Common European Framework (CEFR), and results are reported using the six levels of the CEFR. Alderson (2005) further observes that feedback offered in diagnostic language tests might help learners to

progress from their current level to the next level up on the CEFR.

As Alderson (2005) indicates, the assessment of grammatical abilities in DIALANG is not unlike normal practice in foreign language testing. However, it still focuses on the use of the language (Alderson, 2005). Grammar tests in DIALANG are largely sentence-based and less conservative than they are often thought to be than tests of other skills. Since there is not much empirical evidence regarding how foreign language learners develop their grammatical competence, the DIALANG creators began to develop a set of self-assessment statements which might eventually be used to inform test specifications and contribute to the development of self-assessment of grammatical ability. As for the construct of grammar in DIALANG, the items in the grammar tests measure the test-taker's ability to 1) understand and use morphology (e.g. nouns, articles, verbs, etc., and 2) understand and use syntax (word order, questions, punctuation, etc.).

2.3.2 Self-assessment and Diagnostic Language Testing

Students need to develop the capacity for self-assessment so that they can learn to "notice" for themselves how their language compares with the target-language norms (Purpura, 2004). Further, the learner's self-assessment can trigger further reanalysis and restructuring of their inter-language, and can foster the development of skills required to regulate their own learning (Rief, 1990; Purpura, 2004).

Self-assessment takes place when learners judge and adjust their ability after receiving feedback. This process includes internalizing the explicit knowledge and the facilitation of learner autonomy and self-assessment is an integral part of the feedback system (Alderson, 2005). Even pre-school children can identify certain criteria for good work (Sperling, 1993; Purpura, 2004). Therefore, self-assessment happens in learner learning process at an early age.

Self-assessment is the encouragement of learner autonomy (Brown & Hudson, 1998). In addition, the principle of autonomy stands out as one of the primary foundation stones of successful learning and he goes on to say that learner autonomy helps in developing learners' intrinsic motivation (Brown, 2010).

In DIALANG project, Alderson (2005) defined self-diagnosis as learner reflection upon their performance after feedback is given, and comparison between their belief and expectations about their learning ability. He further states that self-assessment is believed to be central to language learning and a central component of DIALANG. Learners are offered the opportunity to connect their inter-language and information elicited from the feedback.

Diagnostic tests could be extremely useful for individualized instruction or self-instruction (Hughes, 1989; Purpura, 2004). The DIALANG project has demonstrated that integrating self-assessment into diagnosis can facilitate students' involvement in the

assessment process, enhance their metacognitive ability to evaluate their learning outcomes, monitor their own progress, and adjust their own learning (Alderson, 2005; Jang, 2009).

2.3.3 Diagnostic Grammar Testing

Diagnostic grammar tests provide teachers and learners with concrete information. The diagnostic grammar test feedback provides information on what aspects of the grammar students have or have not mastered. This information enables both students and teachers to know what to work on in the future and permits learning to take place independently (Purpura, 2004). He goes on to say that diagnostic grammar testing is a kind of learning-oriented assessment that is not only concerned with issues of grammar testing, but also with issues of instructed learning.

"Learning-oriented assessment aims to provide information that students know, understand, or can use in certain contexts, and the implications that this information might have for grammar processing" (Purpura, 2004, p. 216). "Finally, moving beyond grammar performance per se, learning-oriented assessment can also provide teachers with information about what students feel or believe about learning grammar and about themselves as learners of grammar" (Purpura, 2004, p. 216).

Studies related to computer-assisted grammar testing have been conducted by

several researchers. Nagata and Swisher (1995) found that computers can raise grammatical awareness by giving intelligent feedback (a detailed explanation) to students (Heift, 2001). Nutta (1998) examined whether computer-based grammar instruction is as effective as teacher-directed grammar instruction for postsecondary students at multiple levels of proficiency in an intensive ESL program. Nutta's results show that for all levels of English proficiency, the computer-based students scored significantly higher on open-ended tests covering the structures in question than the teacher-centered students. Further, no significant differences were found between the computer-based and teacher-directed students' scores on multiple choice or fill-in-the-blank test.

Diagnostic grammar test feedback is usually explicit and immediate, not implicit or delayed. With the advent of computer-assisted grammar testing, diagnostic grammar test feedback is provided easily and immediately to help learners monitor their performance and increase their grammatical knowledge. Nagata (1996) investigated the effectiveness of two different levels of computer feedback for teaching Japanese passive sentences. The results show that feedback with detailed grammatical explanations is more effective than feedback with only missing/unexpected word in the learners' achievement in producing Japanese passive sentences.

In addition, in diagnostic language testing, self-assessment plays an important role in the learner's learning process when they internalize the feedback information after taking

diagnostic language tests (Alderson, 2005). Students can learn to "notice" for themselves how their inter-language compares with the target language norms. At the same time, learner autonomy is enhanced, maximizing the efficiency of self-assessment and optimally leads to facilitative learning (Purpura, 2004). Furthermore, the computer-assisted diagnostic grammar tests could be applied in EFL settings, and with the provision of diagnostic feedback, diagnostic grammar tests become the facilitation of effective teaching and learning.

Even though several empirical studies support the usefulness of grammar test feedback, perceptions of usefulness do not necessarily mean learning has occurred. Yin *et al* (in press) point out that the value of a diagnostic language test depends on how useful the feedback is to test-takers. However, the concept of "usefulness" is rather subjective, being based on students' individual perceptions of relevance and meaningfulness. Therefore, the result of any attempt to measure the usefulness of feedback should be examined with caution. Whether the feedback actually leads to effective learning of grammar requires investigation, especially with diagnostic tests that aim to address weaknesses. Regrettably, such an investigation was beyond the scope of this study.

2.4 Feedback

2.4.1 The Value of Feedback

An important feature of feedback is that it allows learners to relate the information to any other materials in their world to improve their language ability. Further, test-takers can explore advice on how they can improve from their current level to higher levels (Alderson, 2005).

Feedback is the provision of meaningful information to the test-taker and is the essence of a diagnostic test (Alderson, 2005). Further, self-assessment is a crucial part of DIALANG, and an integral part of the feedback system (Alderson, 2005). Self-assessment includes internalization of the explicit knowledge and facilitation of learner autonomy. Furthermore, feedback aims at informing learners, supporting learning and raising awareness, and optimally leads to enhancement of learners' language level.

L2 learners need to become aware of the input before they begin to acquire a language feature, "noticing" is the essential starting point, and is vital for L2 inter-language development (Schmidt, 1990). Awareness at the level of noticing is required for learning (Schmidt, 2001). Providing learners with feedback helps them compare or "notice" the differences between their inter-language of the target language (Purpura, 2004). Further, Schmidt (1990, 1993) and Sharwood Smith (1993) claim that feedback makes learners be able to accommodate the differences between their

inter-language and target language, thereby contributing to the ultimate internalization of the learning point (Purpura, 2004).

Table 2.2
Explicit Feedback & Implicit Feedback

Explicit Feedback	**Implicit Feedback**
1.explicit rejection	1.recasts
2.explicit correction	2.requests for clarification
3.metalinguistic information	

2.4.2 Types of Feedback

2.4.2.1 Explicit Feedback V.S. Implicit Feedback

"Merely presenting users with a test score, without explanations, is not very helpful" (Alderson, 2005, p. 208). Explicit feedback includes explicit rejection, explicit correction and metalinguistic information (Ellis, 2008). He goes on to say that there are two forms of explicit feedback: one is explicit correction, which provides both positive and negative evidence; the other is metalinguistic feedback.

As for explicit correction, Lyster and Randa (1997) defined it as "the explicit provision of the correct form" (cited in Ellis, 2008, p. 442). They further state that "metalinguistic feedback contains comments, information, or questions related to the well-formedness of the student's utterance, without explicitly providing the correct form" (Ellis, 2008, P. 442). Purpura (2004) also points out that metalinguistic feedback involves

the use of linguistic terminology to promote "noticing", which, helps the input become intake and optimally leads to facilitative learning.

Implicit feedback refers to recasts and requests for clarification (Ellis, 2008). Ellis *et al* (2006) also argue that implicit feedback often takes the form of recast, which is defined by Long (1996) as "a formulation of all or part of a learner's immediately preceding utterance in which one of more non-target like (lexical, grammatical etc.) items are replaced by the corresponding target language forms(s), and where, throughout the exchange, the focus of the interlocutors is on meaning not language as an object" (Ellis *et al*, 2006, p. 340).

Recasts provide positive evidence, which offers the correct form of the error (Ellis *et al*, 2006). Nicholas, Lightbown, and Spada (2001) also point out that it is not clear that recasts provide negative evidence, which gives exact location of the error and makes learners have conscious awareness that the recasts are intended to be corrective (Ellis *et al*, 2006). Recasts connect linguistic form to meaning in discourse contexts that promote noticing required for implicit language learning (Long, 1996; Doughty, 2001; Ellis *et al*, 2006).

Doughty (2001) further indicates that explicit feedback, such as metalinguistic feedback, is more likely to result in explicit rather than implicit second language knowledge. However, Ellis *et al* (2006) point out that Long (1996) and Doughty (2001)'s

assumption is not unproblematic because it is not certain that all recasts are implicit.

In addition to these definitions, research has been conducted on the effectiveness of explicit feedback and implicit feedback. Heift (2001) discusses learners' responses to metalinguistic feedback and their strategies in error correction in a web-based Intelligent Language System (ILTS) for German. The results indicate that for the majority of sentences (79.5%) students read and attended to system feedback.

Heift (2004) investigated the effects of corrective feedback on learner uptake in CALL by examining 177 Canadian university students' responses in three types of corrective feedback. The feedback provided differed in 1) the amount and 2) specificity of information. The study found that the more explicit and prominent the feedback, the more likely students will revise their errors in written grammar and vocabulary exercises.

Carroll, Roberge, and Swain (1992) examined the effectiveness of explicit corrective feedback on two complex French non-suffixes and found that the explicit corrective feedback group outperformed the no feedback group (Ellis *et al*, 2006). Carrol and Swain (1993) found that metalinguistic feedback is more helpful to adult second language learners (Heift, 2001).

Ellis *et al* (2006) investigated the effect of two types of online corrective feedback on the acquisition of past tense –ed by low-intermediate ESL students in Auckland, New Zealand. The study used an experimental design (two experimental groups and a control

group). The participants completed two communicative tasks during which they received with recast (implicit feedback) or metalinguistic explanation (explicit feedback) in response to any utterance that contained an error in the target structure. The results showed that learners' performance on the posttests explicit feedback (metalinguistic explanation) was more effective than implicit corrective feedback (recast).

Bitchener (2008) presents the results of a 2-month study of the efficacy of written corrective feedback to 75 low-intermediate international ESL students in Auckland, New Zealand. The study included three treatment groups: direct corrective feedback, written metalinguistic explanation and oral metalinguistic explanation, plus one control group. The study found that the accuracy of students who received written corrective feedback in the immediate posttest was greater than that of students in the control group and that this level of performance was retained 2 months later.

Sheen (2004) found that repair occurred less frequently following recasts that following explicit correction and metalinguistic feedback in four different instructional contexts (Ellis *et al*, 2006).

In sum, feedback helps learners to connect the target language structure to their inter-language by making learners "notice" (Purpura, 2004). Therefore, the Noticing Hypothesis is a key rationale behind using feedback to improve learning. Further, Heift (2004), Carroll, Roberge, and Swain (1992), Ellis *et al* (2006) and Bitchener (2008) all

give evidence that explicit feedback is more effective than implicit feedback.

2.4.2.2 Immediate Feedback vs. Delayed Feedback

Traditionally, feedback is provided every few weeks (school exams) or months (high-stake tests like TOFEL) after the tests. However, there is little value in this kind of delayed feedback because test-takers will inevitably have forgotten how they performed and why they responded the way they did (Alderson, 2005). Therefore, Alderson (2005) says that "feedback must be given as soon as possible, or the test takers would inevitably forget how they performed, and why they respond" (p. 208). Hence, with the assistance of computer technology, learners can receive immediate feedback on their performance taking language tests. Immediate feedback has maximum impact on the learners' development of inter-language because they can still recall their reasons for responding the way they did, and are more receptive to the feedback. Thus, feedback becomes maximally informative and relevant (Alderson, 2005). Further, if learners can receive feedback immediately, and reflect on their inter-language, incorporation of assessment and language learning becomes urgent (Alderson, 2005).

Nagata (1996) described an intelligent CALI system called "Nihongo-CALI" (Japanese Computer Assisted Language Instruction) that provides immediate, grammatically, sophisticated feedback to students in an interactive environment. She

compared the effectiveness of the two different levels of computer feedback for teaching Japanese passive sentences. The two different levels of computer feedback are traditional feedback, which indicates only missing/unexpected words; and intelligent feedback, which provides detailed grammatical explanations. The study found that intelligent feedback is more effective than even the enhanced version of traditional feedback in enhancing learner ability to produce Japanese passive sentences.

Table 2.3
Immediate Feedback & Delayed Feedback

Immediate Feedback	Delayed Feedback
1) With the assistance of computer technology, learners can receive immediate feedback on their performance taking language tests.	1) Feedback is provided every few weeks (school exams) or months (high-stake tests like TOFEL) after the tests.
2) Immediate feedback has maximum impact on the learners' development of inter-language because they can still recall their reasons for responding the way they did, and are more receptive to the feedback.	2) Test-takers will inevitably have forgotten how they performed and why they responded the way they did.

2.4.3 Diagnostic Feedback

The core of diagnostic testing is feedback, which is provided to help learners adjust their current understanding of the target knowledge (Alderson, 2005). According to

Alderson (2005), in DIALANG, diagnostic feedback is in essence advice on how learners might progress from their current level to the next level up on the CEFR. (See 2.3 Diagnostic Language Testing)

Diagnostic language testing's provision of meaningful feedback was defined by Heift (2003) as "response that provides a learning opportunity for students" (Cotos & Pendar, 2008, p. 533). A diagnostic test should be oriented towards learning by providing students with feedback to be acted upon in addition to displaying immediate results; Further, diagnostic tests may enhance learning opportunity by allowing learners to act upon the received feedback, thus make diagnostic assessment interactive (Cotos & Pender, 2008).

Effective diagnostic feedback usually contains features of explicit feedback, and immediate feedback. However, with the exception of Jang (2007), little research has been done on the effectiveness of diagnostic feedback.

Jang (2007) examined the usefulness of diagnostic feedback from the user's perspective in a cognitive diagnostic assessment (CDA) of L2 reading comprehension ability. She found that CDA provides formative diagnostic information about an individual learner's strengths and weaknesses in reading ability.

Research has also been conducted on the effectiveness of different types of computer or written feedback. Even though those studies did not examine diagnostic feedback, they

are still worth mentioning here because they discuss features that are included in diagnostic feedback.

Nagata (1995, 1996) found that an NLP-based intelligent feedback is more effective than traditional feedback in helping students improve their performance.

Nagata & Swisher (1995) investigated the effectiveness of two types of computer feedback: traditional computer feedback that indicates only missing or unexpected words in the learner's response, and intelligent computer feedback that provides metalinguistic explanations of the nature of errors. The study found that computers can raise grammatical awareness by giving intelligent feedback to students, and further improve the learners' grammatical proficiency in the use of complex structures of the target language.

Nagata (1996) compared the effectiveness of the two different levels of computer feedback for teaching Japanese passive sentences. The two different levels of computer feedback are traditional feedback, which indicates only missing/unexpected word; and intelligent feedback, which provides detailed grammatical explanations. The study found that intelligent feedback is more effective than traditional feedback in enhancing learner ability to produce Japanese passive sentences.

Nagata (1997) examined the effectiveness of computer-assisted metalinguistic instruction for teaching Japanese complex grammatical structures. Fourteen second-year Japanese university students participated in the study and were divided into two groups.

One group received metalinguistic feedback after performing exercises on a computer, the other received translation feedback after the same exercises. The study found that the metalinguistic feedback group outperformed the translation feedback group in an achievement test.

Bitchener *et al* (2005) investigated whether the type of feedback given to 53 adult migrant students on three types of error resulted in improved accuracy in new pieces of writing over a 12 week period. The three types of feedback are direct-explicit written feedback, student-researcher 5-minute conferences and direct–explicit written feedback. The study found a significant effect for the combination of written and conference feedback on accuracy levels in the use of the past simple tense and the definite article in new pieces of writing but no overall effect on accuracy improvement for feedback types when the three error categories were considered as a single group.

Bitchener & Knoch (2008) investigated the effectiveness of written corrective feedback on migrant and international student writing. The study included three treatment groups: direct corrective feedback, written metalinguistic explanation and oral metalinguistic explanation, plus one control group. The results indicated that students who received all 3 types of feedback outperformed those who received no feedback. None of the feedback types was any more effective than another. Further, their level of accuracy was retained over 7 weeks.

Table 2.4

Summaries of Previous Studies on Effectiveness of Types of Feedback

Reference	Research Question/Foci	Participants	Method(s)	Relevant Findings
Nagata (1996)	To examine the effectiveness of two different levels of computer feedback (traditional feedback; intelligent feedback) for teaching Japanese passive sentences	18 students in a first-semester Japanese course at the University of San Francisco	Students were divided into 2 groups and use pretest and posttest design.	There is a statistically significant difference between traditional and intelligent feedback (in the learners' achievement in producing Japanese passive sentences), favoring intelligent feedback.
Jang (2007)	To examine the validity argument from the user's perspective by focusing on the usefulness of the diagnostic feedback in reading ability	A. 2,703 test takers took the LanguEdge field tests at 32 domestic and international test sites across 15 countries in 2002 B. 28 students C. 3 teachers (from two TOFEL preparation courses	A. Use test takers' response data from two forms of the LanguEdge RC test for the Fusion Model skill B. Use LanguEdge test takers' self-assessment questionnaire C. Student questionnaires on diagnostic reports	[CDA aims to provide formative diagnostic feedback through fine-grained reporting of test taker's skill mastery profiles.]

Table 2.4

Summaries of Previous Studies on Effectiveness of Types of Feedback (Continue)

Reference	Research Question/Foci	Participants	Method(s)	Relevant Findings
Nutta (1998)	Is computer-based grammar instruction as effective as teacher-directed grammar instruction for teaching L2 structures?	53 students enrolled in an intensive academic ESL institute at a major university in Florida	A. Compare the performance of groups after they received computer-based and teacher-directed instruction	1. For all levels of English proficiency, the computer-based students scored significantly higher on open-ended tests covering the structures in question than the teacher-centered student. 2. No significant differences were found between the computer-based and teacher-directed students' scores on multiple choice or fill-in-the-black test
Heift (2004)	1. What is the distribution of corrective feedback types in relation to learner uptake? 2. What is the distribution of learner uptake in relation to learner variables? 3. How do students rate different types of feedback?	177 students from 3 Canadian universities during the spring semester 2003	Pretest/ posttest and questionnaire	Feedback type has an effect on learner uptake: the more explicit and prominent thee feedback, the more likely students will revise their errors in written grammar and vocabulary exercises, independent of language proficiency and gender.

Table 2.4

Summaries of Previous Studies on Effectiveness of Types of Feedback (Continue)

Reference	Research Question/Foci	Participants	Method(s)	Relevant Findings
Heift (2001)	1. Do students read and attend to meta-linguistic feedback or overlook it? 2. What techniques do students apply in error correction in an ILTS? 3. Do learners believe the system's analysis, or, in the event of an error perform an independent re-analysis?	33 students in two beginner German class	The 33 students spent three one-hour sessions using the Build a Sentence exercises	1. For the vast majority of sentences (79.5%) students read and attend to system feedback. 2. As iterations increased students paid more attention to the feedback messages.
Ellis *et al* (2006)	To investigate the effects of implicit and explicit corrective feedback on the acquisition of past tense-*ed*.	33 low-intermediate learners of second language English in three classes in a private language school in New Zealand	Pretest and posttest design Two experimental groups and a control group Group 1: implicit feedback (recast) Group 2: explicit feedback (metalinguistic explanation) Group 3: control group	There's a clear advantage for explicit feedback over implicit feedback for both the delayed imitation and grammaticality judgment posttests

Table 2.4

Summaries of Previous Studies on Effectiveness of Types of Feedback (Continue)

Reference	Research Question/Foci	Participants	Method(s)	Relevant Findings
Bitchener *et al* (2005)	To investigate whether the type of feedback given to 53 adult migrant students on three types of error (prepositions, the past simple tense, and the definite article) resulted in improved accuracy in new pieces of writing over a 12 week period	53 adult migrant students	Pretest and posttest design *Participants were divided into three treatment groups	There is a significant effect for the combination of written and conference feedback on accuracy levels in the use of the past simple tense and the definite article in new pieces of writing but no overall effect on accuracy improvement for feedback types when the three error categories were considered as a single group.
Bitchener (2008)	To investigate the efficacy of written corrective feedback of a 2-month study	75 low-intermediate international ESL students in Auckland, New Zealand	Pretest, immediate posttest and delayed posttest	Accuracy of students who received written corrective feedback in the immediate posttest outperformed those in the control group and this level of performance was retained 2 months later.
Bitchener & Knoch (2008)	To investigates 1. The efficacy of WCF over time 2. Whether certain WCF options typically used in L2 classrooms are more effective	144 international and migrant ESL students in Auckland, New Zealand	Pretest, posttest and delayed-posttest *3 treatment groups + 1 control group	1. Students who received all 3 WCF options outperformed those who did not received WCF 2. Their level of accuracy was retained over 7 weeks→ None of the feedback options was any more effective than another.

2.5 Research Gap

Diagnostic grammar testing identifies strengths and weaknesses in learners' current knowledge while providing the test-taker feedback that enables them to compare their inter-language and the feedback information (Alderson, 2005). Among the types of feedback, explicit feedback is defined by Ellis (2008) as explicit correction and metalinguistic explanation, and is reported to be more effective than implicit feedback in Carroll, Roberge, and Swain (1992), Ellis *et al* (2006), and Bitchener (2008). Nevertheless, with the exception of Jang (2007), little research has been done on the effectiveness of different types of diagnostic feedback. Most studies merely examined the effectiveness of explicit/implicit feedback or immediate/delayed feedback on assessment administered by means of paper and pencil or computer.

Further, in several studies conducted on the effectiveness of diagnostic grammar test feedback, the usefulness of feedback was based on test-takers' subjective perception and thus lacked independent evidence to support the effectiveness of feedback (Yin *et al*, in press). Therefore, this current study attempts to identify the effectiveness of diagnostic grammar test feedback by investigating the improvement after the students receive different types of feedback.

In addition, most of the studies reviewed in this chapter were conducted in countries including the U.S.A, Canada, Germany and New Zealand, but few studies have been done

on the effectiveness of different types of diagnostic feedback for grammar testing in Asian countries such as Taiwan. Therefore, the current study investigates the effectiveness of different types of online diagnostic feedback on the grammar proficiency of university EFL freshmen in central Taiwan, to be significant.

2.6 Summary

In this chapter, literature, studies and findings relate to this current study are reviewed and discussed. To explore the phases of grammar learning, types of grammatical knowledge, the strong and weak interfaces, the noticing hypothesis and the consciousness-raising are covered in the first section. Schmidt (2001) defines explicit learning as “learning with metalinguistic awareness” while implicit learning refers to “learning without metalinguistic awareness”.

DeKeyser (1998; 2003) argued for a strong interface in which explicit knowledge can be converted into implicit learning. By contrast, Doughty and Williams (1998) and Ellis (2002) argued for a weak interface in which explicit knowledge cannot be directly converted into implicit learning. Further, Schmidt (1990) claims that L2 learners need to become aware of the input before they begin to acquire a language feature and “noticing” is vital for L2 inter-language development. Rutherford (1987) and Schmidt (2001) point out that consciousness-raising (CR) also plays an important role in L2 acquisition and

promotes better implicit learning of grammar. In support of these claims, researchers have done studies to show that consciousness-raising and noticing promote facilitative grammar learning (e.g. Leow, 1997;2000; Mackey, 2006; Izumi, 2002; Sheen, 1992; Fotos ,1993; Sugiharto ,2006).

In assessing grammar knowledge, types of grammar testing, and the value of computer-assisted grammar testing are discussed. Multiple-choice grammar tests require no expert judgment in scoring, can be scored objectively, and are suitable for computer-assisted grammar testing. Hence, multiple-choice grammar questions are adopted in this current study. Purpura (2004) and Alderson (2005) point out that grammatical knowledge is crucial in acquiring a foreign language and Chapelle and Douglas (2006) argue that computer technology makes language testing more efficient by providing immediate feedback to the test-taker.

In addition to the phases of grammar learning and testing, discussion of diagnostic language testing is included in this chapter. According to Alderson (2005), diagnostic language testing identifies strength and weakness in learners' knowledge, offers information for both teachers and students to act upon in the future. He further indicates that feedback is the crucial element in diagnostic tests. Further, with the advent of technology, feedback can be provided immediately and make the learning facilitative (Alderson, 2005; Brown, 2004).

Moreover, after receiving feedback, students self-assess themselves. They notice how their inter-language compares with the target-language norms by internalizing the feedback information after taking diagnostic language tests. At the same time, learner autonomy is enhanced, and leads to facilitative learning (Purpura, 2004).

Explicit feedback is in the form of explicit correction and metalinguistic explanation as for the types of feedback, (Ellis, 2008). Further, Carroll, Roberge, and Swain (1992), Ellis *et al* (2006), and Bitchener (2008) found that explicit feedback is more effective than implicit feedback. Alderson (2005) argues that immediate feedback has maximum impact on the learners' development of inter-language. In support of this argument, Nagata (1996) found that immediate computer feedback is more useful than traditional feedback.

Chapter Three

Methodology

3.1 Participants

The participants were 90 Freshman English for Non Majors (FENM) students (55 females and 35 males) from the Management College at a private university for the 2011 academic year. Participants' gender was not taken into consideration in this study but was provided as background information. At this particular university students are divided into FENM classes based on their English placement test scores. There are three different levels of classes (high, mid, and low). The participants came from one high, one mid, and one low level FENM section. The English skills of the three different levels of FENM classes are reflective of the total FENM population at this university. There are approximately 30 students in each of the FENM class. All FENM students are required to take four hours of English classes each week. FENM at this university is a four skill course, focusing on listening, speaking, reading and writing. Grammar is embedded in the course design and is an essential element of the curriculum.

The three FENM classes were divided into one control group and one treatment group. Students were assigned to groups based on their placement test grammar scores. For each class, students' grammar scores on the placement test were ranked from highest

to lowest, and they were assigned to either the control group or the treatment group on an odd and even basis. This was done to insure that there was no significant difference in grammar ability between the two groups at the onset of the study. A flip of a coin determined which group was the control group and which group was the treatment group.

The participants from the Management College were composed of Accounting, Statistics, Information Management and Finance majors (23 Accounting, 13 Statistics, 14 Information Management and 40 Finance students). These students were non-English majors who have learned English for at least 6 years in middle and high schools. This study took place during the first month of their freshmen year.

3.2 Research Design

3.2.1 Instrument

The two instruments (pre- and post-test) that were used in this study were derived from the grammar section of the Online English Assessment System (OEAS). The OEAS was designed primarily for teachers and students at the research site in central Taiwan to self-assess general English proficiency in three skills, listening, reading, and grammar (see Yin *et al*, in press). This study employed the Grammar section of the OEAS as the pre-test. Questions on the OEAS were in a multiple-choice format. Different types of feedback were provided immediately via the OEAS platform after the students finished

the pre-test. Both the questions and feedback on the pre-test were composed by native English speaking teachers who have over twenty years of teaching experience (Yin *et al*, in press).

In order to make the outcome/ result of the two instruments (pre-and post-test) valid and reliable, procedures suggested by Brown (2004) were followed. The procedures included: 1) determining the purpose of the test, 2) designing test specifications, 3) constructing test items, 4) evaluating and revising test items, 5) specifying scoring procedures, and 6) performing validity and reliability studies (Sims, 2006).

Both the pre-test and post-test were of a similar structure and nature in terms of 1) number of questions, 2) kind of questions, 3) order of questions, and 4) content validity. Each test was composed of 30 multiple-choice questions. The pre-test consisted of 30 questions selected from the OEAS Grammar section based on item analysis done by previous researchers (see Appendix A). The post-test consisted of 30 questions and was considered nearly equivalently to the pre-test (see below).

Each test was designed to measure 15 specific grammar constructs. The 15 constructs were: 1) word order in sentences, 2) subject-verb agreement, 3) coordinated clauses, 4) adverbials, 5) conditionals, 6) relative clauses, 7) noun clauses, 8) past tense, 9) present tense, 10) future tense, 11) modals, 12) comparison, 13) articles, 14) passives, and 15) infinitives and gerunds. According to Yin *et al* (in press), these 15 constructs were

sufficient essential elements in assessing learners' grammar ability.

Both tests had equal numbers of corresponding items and questions with each test, including two questions for each of the 15 constructs listed above. For example, question 1 and 2 were designed to measure *word order in sentences*; questions 3 and 4 were constructed to measure *subject-verb agreement*. Each test had 30 multiple-choice questions with three distractors and one correct answer.

Expert rating as suggested by Chapelle (1999), Alderson *et al* (2002), and Bachman (2004) was used to determine the content validity of both tests. Three FENM teachers who were not involved in the construction of the tests determined that both tests were valid measures of the desired constructs. Thus, this study assumed that both instruments had appropriate content validity. Furthermore, since both tests were constructed using the same test specificity, they had similar content validity. This similar content validity was confirmed by the three test evaluators.

Item analysis was used to make sure that both tests were equivalent in terms of item discrimination and item difficulty. This was done by piloting the test questions. Item analysis from three pilot studies conducted with 129 FENM students (Pilot I with 60 students; Pilot II with 31 students; Pilot III with 38 students) who participated in this study revealed that each question on pre-test and post-test had a similar item difficulty and item discrimination as its corresponding question on post-test. Details of the three

pilots will be described in the next section.

3.2.1.1 The Pilots

Pilot I

The first pilot was done with 60 FENM students from the Management College during the last month of their freshman year. The students took both the pre-test and post-test (in total 60 questions) online. Revisions were made based on the item analysis (Table 3.1) from the first pilot data.

Table 3.1

Item Analysis of Pilot I

Item	Item Difficulty		Item Discriminability	
	Pre-test	**Post-test**	**Pre-test**	**Post-test**
1	0. 65	0.65	0.10	0.28
2	0.75	0.92	0.31	0.32
3	0.72	0.48	0.26	0.41
4	0.42	0.60	0.37	0.19
5	0.47	0.52	0.28	0.38
6	0.78	0.85	0.26	0.45
7	0.70	0.53	0.49	0.45
8	0.70	0.67	0.49	0.46
9	0.50	0.35	0.9	-0.12
10	0.32	0.35	-0.9	0.6
11	0.55	0.50	0.39	0.26
12	0.40	0.45	0.29	0.38
13	0.72	0.43	0.40	0.26
14	0.48	0.67	0.26	0.41
15	0.48	0.53	0.4	0.27
16	0.42	0.57	0.34	0.37
17	0.58	0.62	0.29	0.29
18	0.58	0.45	0.28	0.20
19	0.73	0.72	0.26	0.37
20	0.38	0.30	0.29	0.26
21	0.28	0.22	-0.11	0.11
22	0.45	0.57	0.20	0.30
23	0.47	0.40	0.19	0.13
24	0.60	0.72	0.40	0.50
25	0.50	0.55	0.0	-0.5
26	0.55	0.52	0.26	0.26
27	0.35	0.88	0.23	0.16
28	0.28	0.47	0.44	0.24
29	0.55	0.42	0.17	0.27
30	0.68	0.87	0.52	0.10

As Table 3.1 shows, for example, the item difficulty of item 3 in pre-test is 72 %, While the item difficulty of the corresponding item in post-test is 48%. Apparently, item 3 in the post-test is more difficult than its corresponding item in the pre-test. Therefore, item 3 in post-test was revised and piloted again. As for the revision of the questions, for example, item 3 on the pre-test was:

The old man standing under the park trees _____ happy.

a. do not look
b. not looking happy
c. does not look happy
d. not look happy

Originally, item 3 on the post-test was:

The teacher talking to the students_____strict.

a. looks
b. look
c. is looking
d. looked

However, the item difficulty of the original question #3 in post-test should be higher (meaning the question should be easier in order to have a similar difficulty item to the corresponding item on the pre-test). Therefore, question # 3 was changed to:

The teacher talking to the students_____strict.

a. do not look
b. not looking
c. does not look
d. not look

Item frequency also indicates that some questions needed to be revised (see Appendix B). In total, there were 6 items that needed to be revised: 3, 7, 9, 13, 27, and 30.

Pilot II

The second pilot took place in the summer of the 2010 academic year and was conducted with 31 FENM school students. The six items mentioned above were revised and retested. Item analysis is shown in Table 3.2.

Table 3.2
Item Analysis of Pilot II

Item	**Item Difficulty**		**DS Index**	
	Pre-test	**Post-test**	**Pre-test**	**Pose-test**
3	0.32	0.23	0.37	0.40
7	0.42	0.52	0.42	0.41
9	0.55	0.61	0.53	0.59
13	0.48	0.42	0.61	0.48
27	0.26	0.58	0.39	0.66
30	0.42	0.58	0.64	0.44

Note: DS Index refers to the discriminability

As Table 3.2 shows, some items still needed to be revised again. Therefore, the researcher decided to revise and pilot these six items again. For example, item difficulty of item 27 in the pre-test and post-test are not similar. Therefore, revisions were made. Item 27 on the pre-test was:

During the earthquake, Mr. Peterson _____ when a bookcase fell down on him.

a .hurt

b .hurted

c .was hurt

d .was hurted

Originally, item 27 on the post-test was:

The roof of the building ____in a storm a few days ago.

a. damaged

b. was damaged

c. were damaged

d. damage

After the second pilot, item 27 on the post-test was changed to:

The window of the building _____in a storm a few days ago.

a. broke

b. breaked

c. was broken

d. was breaked

Pilot III

The third pilot was done with 38 FENM students during the first month of their freshman year in the 2011 academic year. The 6 items (Item 3, 7, 9, 13, 27, and 30) were piloted again, and item analysis revealed that the revised questions had similar difficulty and discrimination on their corresponding item on the post-test (see Table 3.3).

Table 3.3

Item Analysis of Pilot III

Item	Item Difficulty		DS Index	
	Pre-test	Post-test	Pre-test	Pose-test
3	0.84	0.79	0.34	0.47
7	0.53	0.79	0.52	0.39
9	0.45	0.47	0.44	0.47
13	0.55	0.61	0.54	0.51
27	0.42	0.87	0.40	0.23
30	0.71	0.92	0.32	0.25

Note: DS Index refers to the discriminability

In addition to the item analysis from the three pilots, the revision of questions also was based on Item frequency of the 6 revised items as shown in Appendix C. After some revisions were made based on item analysis and item frequency of the three pilots, the post-test was assumed to be nearly equivalent to the pre-test.

3.2.1.2 Reliability

In addition to validity, reliability is another essential element in constructing instruments. According to Brown (2010), a reliable test is consistent and dependable. He goes on to say that, in multiple-choice type tests, the items need to be evenly difficult well distributed and the distractors need to be well designed to make the test reliable.

A split-half reliability coefficient was calculated in order to support the reliability of

the two instruments. As suggested by Bachman (2004), in split-half estimates, a commonly used procedure is to split the test into halves by including the odd-numbered items in one half and the even-numbered items in the other. In this study, splitting the test into halves on an odd-even basis was compatible with the constructs of the two instruments because each item had its corresponding item on both the pre-test and post-test. For example, items 1 and 2 examined the word order in sentences, while items 3 and 4 measured subject-verb agreement.

The reliability of a test could be higher than correlation of its halves (Bachman, 2004). In order to correct the correlation for length, Spearman-Brown split-half reliability estimate was calculated. The collected data was used to check the reliability of the two instruments.

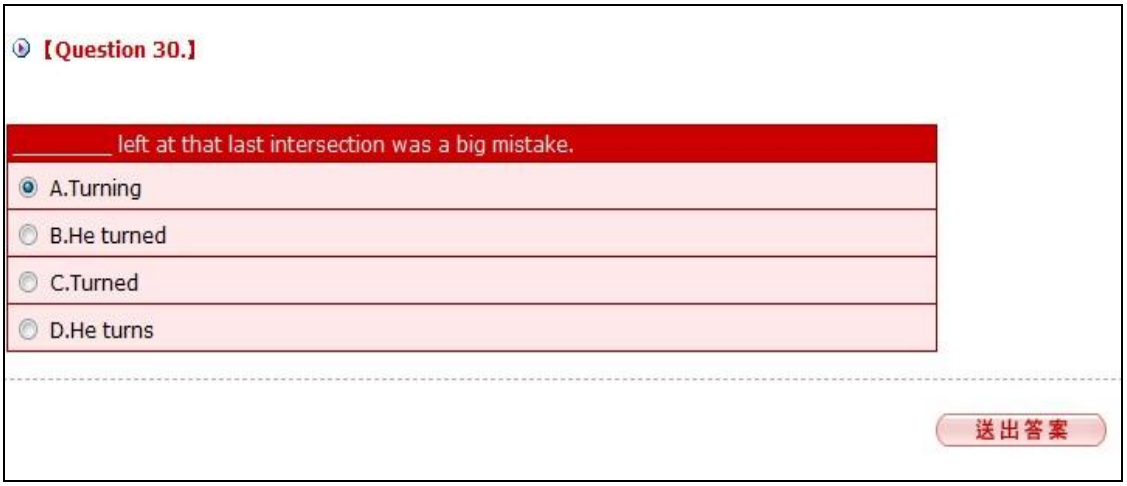
3.2.2 Treatment

Metalinguistic feedback was the treatment used in the current study (see Chapter 2). After the participants took the pre-test, control group students was given only the correct answers without any corrective feedback. However, treatment group students were provided with the feedback (metalinguistic explanations, examples). The procedures for taking the online multiple-choice diagnostic grammar test were under the researcher's instructions and were described as follows. In step one, the test-taker read the questions

and chose one answer in each question (Figure 3.1). After completing the questions, the test-taker clicked the button “send” (see P.49) to submit the answers (Figure 3.2).

Figure 3.1 Procedure for taking the online multiple-choice grammar test (1)

Figure 3.2 Procedure for taking the online multiple-choice grammar test (2)



After submitting the answers, the test-taker was presented with an overall report of the 15 grammar constructs tested (Figure 3.3). The overall report indicated both how

many items the test-takers answered correctly and their total scores. Control group students were only allowed to see the total report and they needed to log out the system. However, the treatment group students were required to click the "訂正" (Correction) button to read detailed metalinguistic feedback (explanations, example) in both English and Chinese (Figure 3.4, also see Appendix D). Researcher and the assistant went around and made sure the students were following researcher's instructions.

Figure 3.3 Procedure for taking the online multiple-choice grammar test (3)

題型		答對題數
A. Word Order in Statements, Questions, and Exclamations 直述句、疑問句及感嘆句中的字序	訂 正	2/2
B. Subject-Verb Agreement 主詞-動詞的一致性	訂 正	2/2
C. Coordinated Clauses 對等連接詞子句	訂 正	1/2
D. Adverbials 副詞	訂 正	2/2
E. Conditionals 條件子句	訂 正	0/2
F. Relative Clauses 關係子句	訂 正	2/2
G. Noun Clauses 名詞子句	訂 正	2/2
H. Past Tense 過去時態	訂 正	2/2
I. Present Tense 現在時態	訂 正	2/2
J. Future Tense 未來時態	訂 正	2/2
K. Modals 助動詞	訂 正	2/2
L. Comparison 比較	訂 正	1/2
M. Articles 冠詞	訂 正	2/2
N. Passives 被動式	訂 正	2/2
O. Infinitives and Gerunds 不定詞和動名詞	訂 正	2/2
分數		26
總分		26

Figure 3.4 Feedback from the online multiple-choice grammar test


【Question 6.】

A man dressed in old, dirty clothes came to our door _____ for food.

A.and begging

B.but begging

C.and begged(正解)

D.but begged

解析:

The coordinating conjunction "but" introduces a contrast or an unexpected outcome. If you say, "The man stopped but didn't say a word", the word "but" indicates that the second sentence is unexpected and surprises you. In the sentence above, the man's begging for food does not surprise you, so B and D are not correct. Secondly, when you use a coordinating conjunction like "and", the words in front and behind "and" should have the same form. In other words, if you put an adjective before "and" you should also put an adjective after "and". The verb "came" is in the past tense while "begging" is a present particle of a verb. They are not the same form, so A is not correct. C is correct because the verb "begged" is the same form as "came".

對等連接詞but表示一種相反或者預期之外的結果。假如你說：" The man stopped but didn't say a word"，那麼but這個字暗示" didn't say a word"使你驚訝以及在你的預期之外。題目中" the man's begging for food" 並沒有使你感到驚訝，所以B、D都不正確。其次，當你使用對等連接詞時，句子兩端的結構要一致。換句話說，假如你在and之前使用形容詞，那麼也應該在and之後使用形容詞。本題在對等連接詞and之前使用過去式came，故之後也應該使用過去式begged，故選C。

3.3 Data Collection Procedures

In this study, the participants were required to meet at a computer room to take the online multiple-choice grammar pre-post and post-test. Before the tests, participants were provided with instruction sheets (in both Chinese and English) with the group list on it (see Appendix E-Appendix J). In addition, beforehand the researcher went to the three FENM classes to give instructions and made sure that all the participants were clear about the test procedures.

The pre-test took place during the second month of the 2011 academic year. Before the test began, the researcher went earlier to the computer room to **t**urn on the computers, made sure the online assessment system page was ready and opened and wrote the schedule and instructions on the white board. After the students came in, the researcher

began settle the students in their seats, **e**xplained the procedures to the students (e.g. how much time they had and how many questions were there for them to finish), and gave the students instructions about the online assessment system (e.g. log in, technical issues).

After all preparations were done, the pre-test began. All 90 participants (45 students in each group) had 30 minutes to finish the 30 grammar multiple-choice questions online. Group 1 (control group) students were only allowed to see their total scores, and were not be allowed to read the feedback (metalinguistic explanations, examples) after finishing the test under researcher and the assistant's supervision.

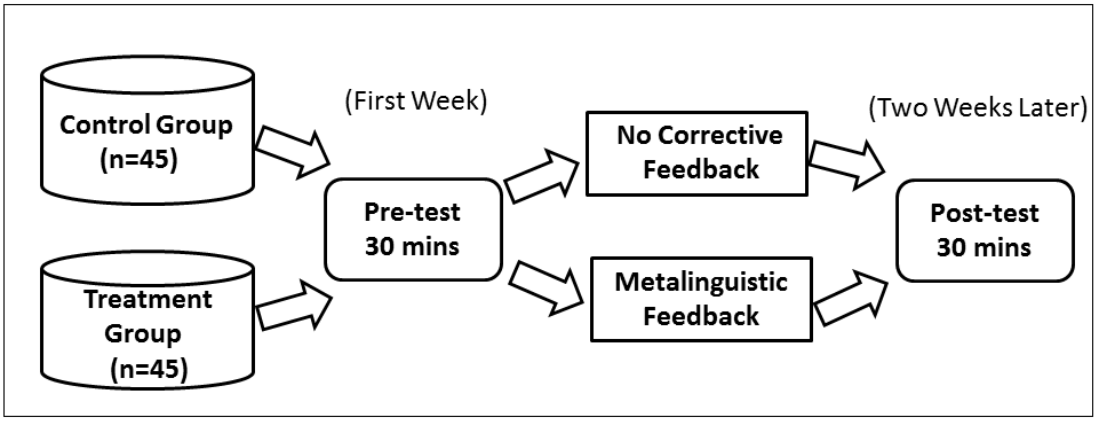
However, Group 2 (treatment group) students saw the right answers to each item, whether their responses were correct or incorrect, and the extended feedback containing metalinguistic explanations, and examples. Further, they were required to write down the time when they started to read the feedback and when they finished reading the feedback on a record sheet (see Appendix K) which was provided by the researcher. A clock was prepared and placed at the front of the class for the students to keep track of the time. Data was collected by the test system automatically.

The post-test took place two weeks after the pre-test. Preparations before the post-test were the same as in the pre-test, including the researcher ensuring that the computers and test system were available, and instructions were given clearly to the students. After all preparations were done, the post-test began. All 90 participants (45

control group students and 45 treatment group students) had 30 minutes to finish the 30 grammar multiple-choice questions online. After the post-test was done, the data was collected automatically by the test system (Figure 3.5).

Figure 3.5

Procedure for the control and the treatment group



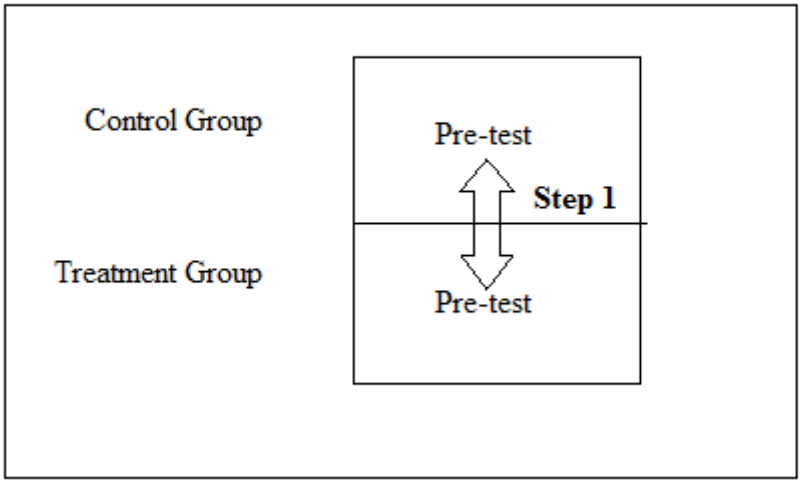
3.4 Data Analysis

The data in this study was analyzed using the SPSS 20.0 statistical package. Two independent-sample t-tests were used to analyze the mean scores of pre-tests and post-tests between groups, while two paired-sample t-tests were used to analyze the mean scores of pre-tests and post-tests within groups.

Data analysis procedures were divided into four steps. The first step was an independent-sample t-test done on pre-tests between the control group and the treatment group (Figure 3.6). Step 2 was a paired-sample t-test conducted on pre-test and post-test within the control group and Step 3 was another paired-sample t-test on pre-test and post-test within the treatment group (Figure 3.7). Finally, Step 4 was an independent-sample t-test practiced on post-tests between groups (Figure 3.8).

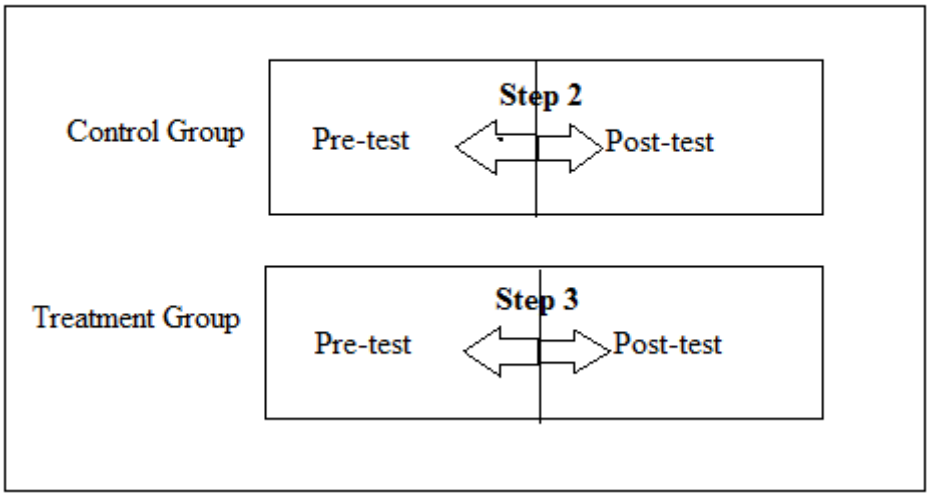
At the onset of the treatment, an independent-sample t-test was done on the pre-test results to make sure there was not a statistically significant difference between the control group and the treatment group (Figure 3.6). This step was set to find out if both the control and treatment group were at the same level of English grammar ability before the treatment (metalinguistic feedback).

Figure 3.6 Data Analysis Procedure Step 1 (Pre-test Between Groups)



After the post-test data was collected, the researcher began to analyze the differences in the scores within and between the two groups. In step 2, one paired-sample t-tests was used to analyze the mean scores on pre-test and post-test within the control group (Figure 3.7). In other words, for the control group, a paired-sample t-test was done between the pre-test and post-test. This step helped the researcher know if the control group students would make any improvement on the post-test even though they did not receive any metalinguistic feedback after taking the pre-test.

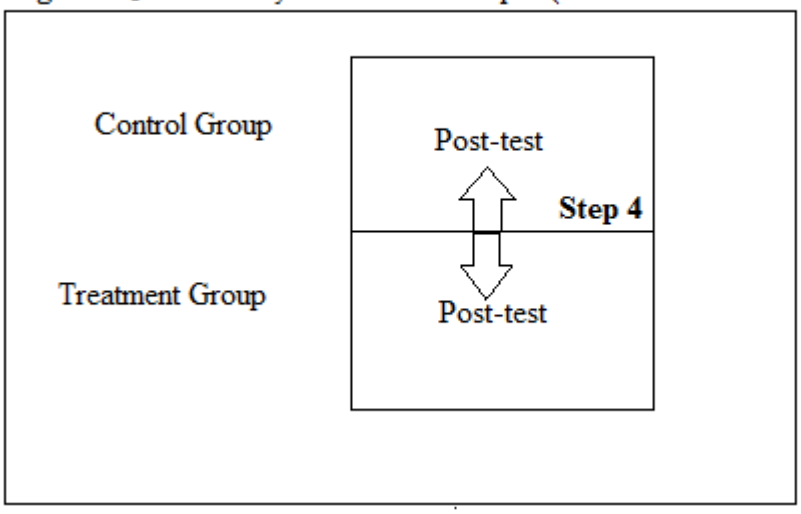
Figure 3.7 Data Analysis Procedure Step 2 and Step 3 (Within Groups)



After the paired-sample t-test done on pre-test and post-test mean scores within the control group in step 2, step 3 was another paired-sampled t-test conducted on pre-test and post-test scores of the treatment group to see if there was a statistically significant difference within the mean scores of the two groups (also see Figure 3.7). This step helped the researcher know whether the treatment group made any improvement on the post-test after receiving the metalinguistic feedback.

Step 4 was another independent-sample t-test that conducted between the control group and treatment group mean scores on the post-test. This determined if the treatment was beneficial by examining whether there were statistically significant differences between the post-test mean scores of each group (Figure 3.8).

Figure 3.8 Data Analysis Procedure Step 4 (Post-test Between Groups)



Methods that were used in this current study were illustrated in this section; detailed results of data analysis are presented in next chapter.

Chapter Four

Results and Discussion

In this chapter, the results of the experiment are presented. There are five sections in this chapter. First, the reliability coefficients of the instruments are examined. Then the reliability results are followed by the statistical results of the research questions. After the investigation of the effectiveness of the treatment (metalinguistic feedback) is conducted, several comparisons of the pre-test and the post-test data between and within groups are reported. In addition, comparisons of the results with that of other studies are discussed. Finally, this chapter ends with a summary of the results.

4.1 Reliability of the Instruments

The reliability coefficients of the pre-test and post-test were obtained using the Spearman-Brown split-half coefficient analysis. A total of 90 participants took the pre-test and post-test. Reliability coefficients of the pre-test and the post-test were calculated based on the data from all the 90 participants. Both the reliability coefficient of the pre-test of 30 test items and of the post-test of 30 test items were over .70 (Table 4.1). Therefore, the reliability of the two instruments was acceptable.

Table 4.1

Split-half Reliability Coefficients of the Pre-test and the Post-test

	Pre-test	**Post-test**
Split-Half Coefficient	0.71	0.73

4.2 Descriptive Statistics of the Tests

This section presents the descriptive statistics of both the pre-test and the post-test. Section 4.2.1 reports the item analysis. Section 4.2.2 presents the mean scores and standard deviation.

4.2.1 Item Analysis

The range of difficulty level in the pre-test and the post-test indicates that the two instruments contained questions ranging from easier to more difficult. In the pre-test, difficulty level ranged from 0.26 to 0.86. In the post-test, difficulty level ranged from 0.27 to 0.97.

In addition to the range of difficulty, the difficulty level of most of the items in the pre-test was similar to its corresponding item in the post-test. The items in the pre-test and the post-test were equivalent to each other. The difference of each item in the pre-test and its corresponding item in the post-test ranged from 0 to 0.30 (Table 4.2).

Table 4.2

Item Difficulty, Pre-test and the Post-test

Item	Item Difficulty		
	Pre-test	**Post-test**	**Difference**
1	**0.73**	*0.79*	0.06
2	*0.79*	*0.96*	0.17
3	*0.78*	**0.66**	0.12
4	**0.58**	**0.61**	0.03
5	**0.39**	**0.49**	0.10
6	*0.79*	*0.90*	0.11
7	*0.83*	*0.83*	0.00
8	**0.73**	**0.69**	0.04
9	**0.56**	**0.52**	0.04
10	**0.26**	**0.40**	0.14
11	**0.51**	**0.58**	0.07
12	**0.52**	**0.59**	0.07
13	*0.76*	**0.72**	0.04
14	**0.57**	**0.68**	0.11
15	**0.51**	**0.66**	0.15
16	**0.58**	**0.72**	0.14
17	**0.51**	**0.60**	0.09
18	**0.72**	**0.61**	0.11
19	*0.86*	*0.88*	0.02
20	**0.28**	**0.27**	0.01
21	**0.29**	**0.31**	0.02
22	**0.42**	**0.56**	0.14
23	**0.46**	**0.47**	0.01
24	**0.70**	*0.76*	0.06
25	**0.51**	**0.63**	0.12
26	**0.53**	**0.60**	0.07
27	**0.56**	*0.86*	0.30
28	**0.34**	**0.59**	0.25
29	**0.61**	**0.42**	0.19
30	*0.83*	*0.97*	0.14

According to *Item Analysis*, if the item difficulty level exceeds .75, it is an easy item; if the level is below .25, it is a difficult item. Thus, item difficulty levels ranging from .75 to .25 are of medium difficulty. In the pre-test, 7 items were easy questions while in the post-test, 8 were considered easy (as marked in italics in Table 4.2). Meanwhile, in the pre-test, 23 items were difficult while in the post-test, 22 were identified as challenging questions (marked in bold face).

Test items of easy and medium level are reported and discussed as follows. Item No. 30 on the pre-test tested students' ability to use gerunds. It tests whether the students knew that adding *ing* to the verb *turn* can change it into a noun. It was an easy question for the participants, 83% of them chose the correct answer.

_____ *left at that last intersection was a big mistake*

a . *Turning*
b . *He turned*
c . *Turned*
d . *He turns*

While item No. 15 on the pre-test tested students' ability to use past tense. It tests whether the students knew that when something happened in the past, the other thing happened before should be past perfect. It was a medium question for the participants, 51 % of them chose the correct answer.

Charlie discovered that he _____ the wrong girl in the dark when the lights went on again.

a. had kissed

b. kissed

c. is kissing

d. was kissed

In addition to the item difficulty, item discriminability of the pre-test and the post-test was obtained by comparing the upper 1/3 and the lower 1/3. Discriminability on pre-test ranged from -0.07 to 0.57, while on post-test, it ranged from 0.07 to 0.47. It should be noted that there was one negative discriminability found among the test items, and the other items had positive values (Table 4.3).

Table 4.3

Item Discriminability of the Pre-test and the Post-test

Item	Item Discriminability	
	Pre-test	**Post-test**
1	0.17	0.27
2	0.47	0.10
3	0.33	0.07
4	0.53	0.37
5	0.37	0.40
6	0.20	0.13
7	0.23	0.20
8	0.27	0.33
9	0.23	0.17
10	0.20	0.10
11	0.40	0.47
12	0.37	0.30
13	0.27	0.47
14	0.17	0.17
15	0.33	0.33
16	0.53	0.30
17	0.30	0.30
18	0.13	0.37
19	0.13	0.27
20	0.07	0.40
21	0.37	0.30
22	**-0.07**	0.40
23	0.40	0.20
24	0.33	0.33
25	0.30	0.33
26	0.27	0.27
27	0.13	0.13
28	0.27	0.23
29	0.57	0.17
30	0.30	0.07

Negative discriminability was found in item No. 22 on the pre-test, while the corresponding item on the post-test exhibited positive discriminability. Item No. 22 on the pre-test and the post-test tested students' ability to identify modals and past tense. Samples of item No. 22 on the pre-test and the post-test were provided as follows.

Sample of Item No. 22 on the pre-test

After having asked the coach for many weeks, I _____ the team yesterday.

a. was finally able to join

b.could finally join

c. could have finally joined

d.can finally join

Sample of Item No.22 on the post-test

After saving money for a long time, I _____the camera yesterday.

a. should finally buy

b. was finally able to buy

c. could have finally bought

d. can finally buy

Table 4.4 indicates that most of the students chose the correct answer "A" in item No. 22 on the pre-test. However, many students chose "B" as the correct answer. One possible reason for the error could be that students choosing "B" as the correct answer did not understand the difference between *can* and *be able to*. According to the metalinguistic explanation, *be able to* means the person has the ability to accomplish something while *can* refers to a possibility and ability of doing something. But when somebody wants to express that something is very hard to achieve, it would be more suitable to use *be able to*. Therefore, students who chose *could finally join* instead of *was finally able to join* in item

No. 22 on the pre-test may have considered *could* as *be able to* in past tense.

Because the pre-test questions were adopted from the OEAS system, they could not be changed. However, item No. 22 on the post-test has been revised and now has positive discriminability. As for item No. 22 on the post-test, *could* has been replaced by *should*. Therefore, most of the students chose the correct answer *was finally able to buy* and were not distracted by the other three choices in this question.

Table 4.4

Frequency (%) of Item No. 22 on the Pre-test and the Post-test

	Pre-test					Post-test				
Item	Key	A	B	C	D	Key	A	B	C	D
22	A	**38.0**	31.0	16.0	5.0	B	2.0	**50.0**	30.0	8.0

4.2.2 Mean Score and Standard Deviation

The mean scores of Group 1 and Group 2 on pre-test were both near 18.0, and the standard deviations of the two groups were near 4.0 (Table 4.5).

Table 4.5
Mean Scores and Standard Deviation

	Pre-test			Post-test		
	N	**M**	**SD**	**N**	**M**	**SD**
Group 1	45	17.53	4.17	45	18.18	3.50
Group 2	45	17.72	3.68	45	20.42	3.50

Mean scores of Group 1 and Group 2 on the pre-test indicated that the two groups had similar ability at the onset of the study. The mean score on the post-test for Group 1 was near 18.0 while for Group 2 it was near 21.0. The mean scores of the two groups on the post-test indicated that Group 2 students performed better than Group 1 students on the post-test. However, whether the difference between the mean scores of Group 1 and Group 2 on the post-test was statistically significant requires statistical analysis. Detailed results are reported in 4.3. The test scores of each participant are given in Appendix M, and the frequency of each item on the pre-test and the post-test are given in Appendix N.

In addition to the mean score and standard deviation, the score distribution of the pre-test and the post-test in Group 1 and Group 2 is documented. Test distribution of the pre-test and post-test was close to the normal distribution (Figures 4.1 to 4.4). Most

scores fell between the interval of 16 and 20 (Table 4.6).

Table 4.6

Frequency of the Pre-test and the Post-test in Group 1 and Group 2

	Group 1		Group 2	
Scores	**Pre-test**	**Post-test**	**Pre-test**	**Post-test**
1~5	0	0	0	0
6~10	3	1	1	0
11~15	10	7	12	3
16~20	20	25	21	19
21~25	12	11	11	21
26~30	0	1	0	2
N	45	45	45	45

Figure 4.1

Frequency of Pre-test in Group 1 (N=45)

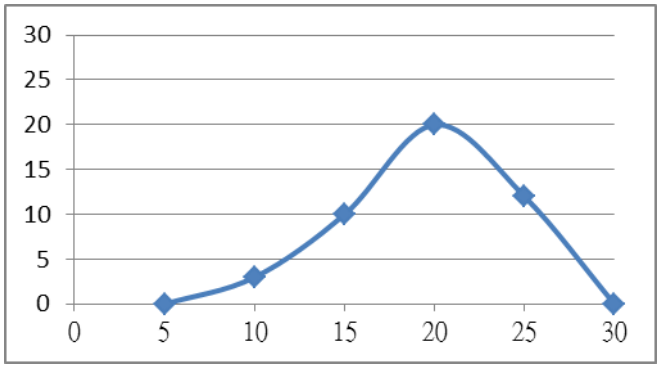


Figure 4.2

Frequency of Post-test in Group 1 (N=45)

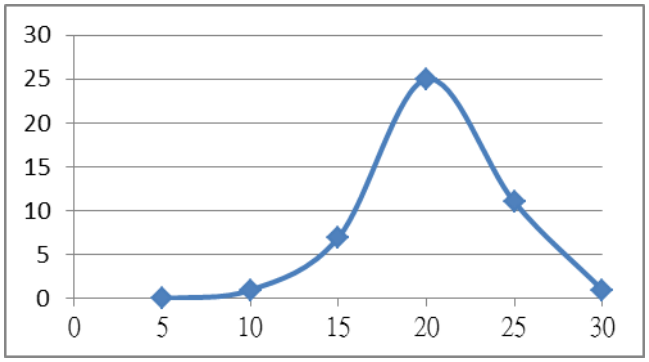


Figure 4.3

Frequency of Pre-test in Group 2 (N=45)

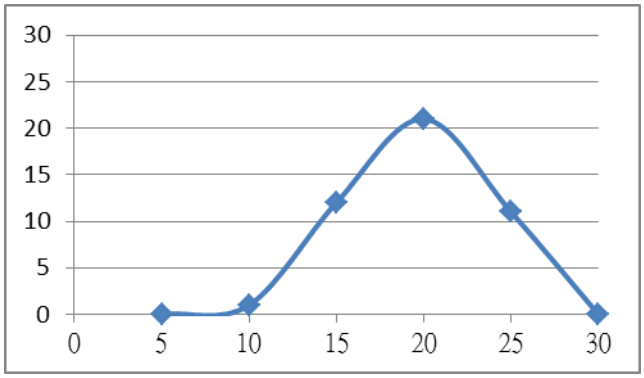
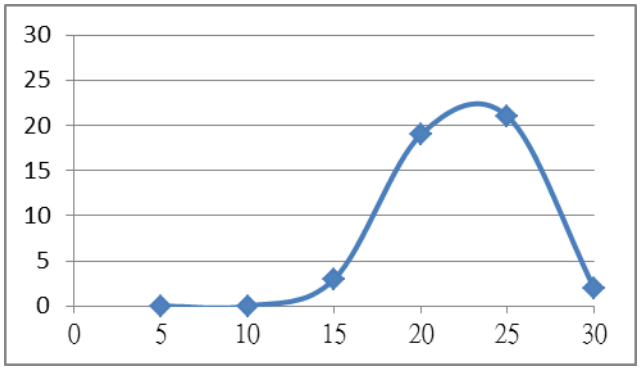


Figure 4.4

Frequency of Post-test in Group 2 (N=45)



4.3 Comparison of Participants' Performance on the Pre-test and the Post-test

This section contains three parts, presented in the order of the research questions. Major findings and answers to the research questions are presented. The first part (4.3.1) presents the pre-test results. The second part (4.3.2) provides the results of the analysis of the mean score differences in the pre-test and post-test within the two groups. The last section (4.3.3) reports the post-test results.

4.3.1 Pre-Test Results

An independent-sample t-test indicated that both the control group (Group 1) and treatment group (Group 2) did not differ significantly in their mean scores on the pre-test. The mean difference was 0.19. In other words, levels of grammar ability of control group and treatment group students were similar at the start of the study (Table 4.7).

Table 4.7
Independent-sample T-test Results of Group 1 and 2 Pre-test Performance

Source	**Group 1 (n=45)**		**Group 2 (n=45)**		***t***	***p***	**95%CI**		**η2**	**1- ß**
	M	**SD**	**M**	**SD**			***LL***	***UL***		
Total	17.53	4.17	17.72	3.68	.134	*p*>.05	-1.54	1.76	.00	.052

**df*=88

The p-value indicated that there was no statistically significant difference between the pre-test results of both the control and treatment group. Further, the confidence

interval showed no significance difference in pre-test performance between the groups because the range of the lower and the upper contained 0.

Since the t-test results revealed that both the control and treatment groups were at a similar level of English grammar ability on the pre-test, the following procedures, including Group 2 students reading the metalinguistic feedback and both groups taking the post-test, were accomplished.

4.3.2 Pre-test and Post-test Results Within Groups

This section contains two parts. First, the results of a paired-sample t-test conducted on mean scores of pre-test and post-test in the control group (Group 1) are presented. Second, the results of another paired-sample t-test done on the mean scores of pre-test and post-test in the treatment group (Group 2) are given.

To know whether the treatment was effective in helping participants improve their grammar ability, scores of the participants of both groups on the pre-test and the post-test needed to be calculated.

Group 1 students received no metalinguistic feedback (metalinguistic explanations, examples) after taking the pre-test. In other words, Group 1 students took the post-test directly, without receiving the treatment.

The results revealed that there was no statistically significant difference between

means scores of the pre-test and the post-test of Group 1. Since the p-value is >.05, the difference between mean scores of pre-test and post-test in Group 1 was not statistically significant. In addition to the p-value, the confidence interval showed that there was no statistically significant difference between the pre- and the post- tests mean scores in Group 1 because the range between the lower and the upper bound included 0 (Table 4.8).

Table 4.8

Paired-sample T-test Results of Performance of Group 1, Pre-test and Post-test

Source	**Pre-test (n=45)**		**Post-test (n=45)**		***t***	***p***	**95%CI**	
	M	**SD**	**M**	**SD**			***LL***	***UL***
Total	17.53	4.17	18.18	3.49	-1.29	*p*>.05	-1.65	.36

**df*=44

Unlike the Group 1 students who did not receive any treatment, Group 2 students were informed of the metalinguistic feedback (explanations, examples) immediately after they took the pre-test. Mean scores show that scores of participants in Group 2 increased significantly between pre- and post-testing.

In addition, the p-value indicated that the difference between the mean scores of the pre-test and the post-test in Group 2 was statistically significant. The confidence interval was also significant. The 2.95 point range between mean scores of the pre-test and the post-test showed that participants in the treatment group made a substantial improvement on the post-test (Table 4.9).

Table 4.9

Paired-sample T-test Results of the Performance of Group 2, Pre-test and Post-test

Source	Pre-test (n=45)		Post-test (n=45)		*T*	*p*	95%CI	
	M	SD	M	SD			*LL*	*UL*
Total	17.47	3.73	20.42	3.50	-5.70	*p*<.001	-4.00	-1.91

**df*=44

4.3.3 Post-Test Results between Groups

This section presents the results of an independent-sample t test conducted on the mean scores of post-tests of Group 1 and Group 2. The results indicated that Group 1 and Group 2 did not differ significantly in their mean scores on post-tests. Further, for a p-value under .01, statistically significant differences were not found between the mean scores of Groups 1 and 2 (Table 4.10).

Table 4.10

Independent-sample T-test Results for Performance of Groups 1 and 2 on Post-test

Source	Group 1 (n=45)		Group 2 (n=45)		*t*	*P*	95%CI		η2	1- ß
	M	SD	M	SD			*LL*	*UL*		
Total	18.18	3.50	20.42	3.50	.003	*p*<.01	-3.71	-.78	.10	.854

**df*=88

4.4 Discussion of the Results

Data analysis in the current study was divided into four stages, including two independent-sample t-tests for analyzing the mean scores between the groups and two paired-sample t-tests to analyze the mean scores within the groups. Different measures of analyzing the data were used to answer the research questions.

At the beginning of the study, one independent-sample t-test was performed on the pre-test scores of both the control and treatment groups. The results indicated that there was no significant difference between the pre-test mean scores of the two groups. The lack of significance differences between the two groups at this stage is likely the result of the equal division of the participants by grammar ability into groups based on their FENM placement test scores. When dividing the three levels of FENM students (n=90) into one control and one treatment group, their placement test scores were used as the criteria. Therefore, it could be said that the placement test and instrument used in this study had similar validity.

For the analysis of pre-test and post-test scores within the control group, a paired-sample t-test was used. The results indicated that there was no significant difference between the pre-test and post-test mean scores within the control group. One reason for this may lie in the absence of metalinguistic feedback. In the present study, the control group did not receive any kind of corrective feedback. Participants in this group

were only informed of their total scores. In addition, they were not allowed to read the metalinguistic feedback after the pre-test. Instead, they were required to log out of the online test system immediately. Another reason may be that participants in the control group did not have the opportunity to acquire additional grammatical knowledge in their communicative or educational settings during the interim period between the pre-test and post-test.

In terms of the other paired-sample t-test conducted on the pre-test scores and post-test scores within the treatment group, the results revealed that the treatment group increased significantly between pre-testing and post-testing. A partial explanation for this result may lie in the fact that the treatment group students did receive metalinguistic feedback immediately after they finished the pre-test. Metalinguistic feedback, including explanations and corrections for the questions, served as the treatment in this study. After the treatment group students finished the pre-test, metalinguistic feedback was immediately provided by the online test system. Test takers only had to click the "correction" button to read the explanations and corrections in both English and Chinese for the test questions.

Metalinguistic feedback in this study was believed to be connected with students' interlanguage, further strengthening their grammatical knowledge and leading to better performance. In addition, perhaps the interim period between the pre-test and post-test

allowed the students in treatment group to apply the grammatical structure/knowledge in questions/explanations in real communicative situations, providing them with additional opportunity to negotiate meaning and monitor their own output.

For the last stage of data analysis, an independent-sample t-test of the post-test mean scores of control group and the post-test means scores of treatment group was performed. Results of this stage showed that there were not significant differences in the post-test scores of the two groups though the mean scores of treatment group were higher than those of the control group.

One for this outcome may be that the treatment group received metalinguistic feedback while control group did not. Further, because both control and treatment group were of equal level of grammar proficiency at the start of the study, if treatment group made improvement on the post-test, then metalinguistic feedback had been proved to be efficient. Therefore, it is possible that treatment group could do better than control group on the post-test (Table 4.11).

Table 4.11

Summary of Data Analysis and Results

Group	Data Analysis	Significance
Control (pre-test) & Treatment (pre-test)	Independent-sample t-test	No
Control (pre-test) & Control (post-test)	Paired-sample t-test	No
Treatment (pre-test) & Treatment (post-test)	Paired-sample t-test	Yes
Control (post-test) & Treatment (post-test)	Independent-sample t-test	No

4.5 Comparison between Results and the Literature Review

This section compares the results of the current study with those of other studies that were mentioned in Chapter Two. Type of feedback in previous studies were discussed in 2.4.2. Findings in the current study are not in contradiction with those of the empirical studies mentioned above. With regard to effectiveness of feedback, our findings confirm those of Nagata (1996) and Jang (2007), although there are some differences regarding the aspects of the studies.

In Nagata (1996), the results revealed that intelligent feedback, which provides detailed grammatical explanations, was more effective than traditional feedback, which indicates only missing/unexpected words in enhancing learner ability to produce Japanese

passive sentences.

Also, in Jang (2007), the results indicated that formative diagnostic feedback was useful in enhancing students' L2 reading comprehension ability. Findings in the present study, and those of Nagata (1996) and Jang (2007) all provided evidence that metalinguistic type of feedback (detailed explanations) was useful in helping students improve their second language ability.

In addition, Ellis *et al* (2006) also suggests that learners' performance on the post-test explicit feedback (metalinguistic explanation) was more effective than implicit corrective feedback (recast) on the test of past tense –ed for low-intermediate ESL students. Even though there were some differences between the present study and Ellis *et al* (2006), findings of the two studies indicated that explicit corrective feedback was more useful than implicit kind or no corrective feedback.

Bitchener (2008) and Bitchener & Knoch (2008) examined the efficacy of different types of corrective feedback (direct corrective feedback, written metalinguistic explanation and oral metalinguistic explanation) on writing tasks.

Even though results in Bitchener (2008) and Bitchener & Knoch (2008) showed that none of the three types of feedback was more effective than the others, it also indicated that when students received all these three types of feedback, they outperformed those who did not receive any. This also supports the notion that as long as the feedback is

explicit and includes metalinguistic explanations, it helps learners achieve better learning.

Table 4.12

Comparisons between the Present Study and Other Studies

Studies	Language Target	Type of Feedback	Methods
The present study	English grammar	1.No corrective feedback →Pre-test—Post-test (x) 2.Metalinguistic feedback→Pre-test—Post-test (o)	1.Control & Treatment Group 2.Pre-test/ Post-test design
Nagata (1996)	Japanese passive sentences	1. Traditional feedback →Pre-test—Post-test (x) 2. Intelligent feedback →Pre-test—Post-test (o)	1.Control & Treatment Group 2.Pre-test/ Post-test design
Jang (2007)	English reading comprehension	Formative Diagnostic feedback →Test-taker' perspective (o)	1. Test-taker' response 2. Questionnaire
Ellis *et al* (2006)	English past tense-*ed*.	1.Implicit corrective feedback →Pre-test—Post-test (x) 2.Explicit corrective feedback→Pre-test—Post-test (o)	1.Control & 2 Treatment Group 2.Pre-test/ Post-test design
Bitchener (2008)	English Writing	1. Direct corrective feedback→Pre-test—Post-test (x) 2. Written meta-linguistic explanation→Pre-test—Post-test (o) 3. Oral meta-linguistic explanation→Pre-test—Post-test (x)	Pretest, immediate posttest and delayed posttest
Bitchener & Knoch (2008)	English Writing	1. Direct corrective feedback→Pre-test—Post-test (x) 2. Written meta-linguistic explanation→Pre-test—Post-test (x) 3. Oral meta-linguistic explanation→Pre-test—Post-test (x)	1. 3 treatment groups + 1 control group 2. Pretest, posttest and delayed-posttest

Note: (o) refers to the more effective treatment, and (x) refers to the less effective treatment.

Chapter Five

Conclusion

This chapter firstly presents the conclusion of the present study. The conclusion is followed by implications of the study. Finally, the limitations of the study and suggestions for future research are offered.

5.1 Conclusion

This present study was conducted to examine the effectiveness of different types of feedback (no corrective feedback; metalinguistic feedback) on an online diagnostic multiple-choice grammar test for EFL university freshmen. Investigation of the effectiveness of different types of feedback on students' grammar proficiency helps the researcher determine which types of feedback are the most useful for improving students' grammar ability. Furthermore, the findings of the study are applicable to curriculum design, and will hopefully help facilitate learning in the near future.

Participants (n=90) were from three levels of FENM classes (one high, one mid, and one low) and were divided into one control group (n=45) and one treatment group (n=45). An independent-sample t-test showed that both the control and the treatment group did not differ significantly on the pre-test. After the pre-test, the control group did not receive

any corrective feedback. By contrast, the treatment group received metalinguistic feedback immediately from the online test system. Two weeks after the pre-test, both control and treatment group took the post-test.

Participants took the tests online and data was collected by the computer system automatically. Two independent-sample t-tests were used in analyzing the data between groups, while two paired-sample t-tests were used to analyze the data within the groups. The major findings are summarized as follows.

1) No statistically significant difference was found between the mean scores of the pre-test and post-test within the control group.

2) There were significant differences in the pre-test and post-test mean scores in the treatment group, which received metalinguistic feedback.

3) Differences of post-testing scores between the control and treatment groups were not significant although the mean scores of the treatment group were higher than those of the control group.

The results indicated that treatment (metalinguistic feedback) was the main factor that could explain the differences between the groups in the post-test of the current study. Control group participants did not improve greatly on the post-test, likely because they did not receive any corrective feedback after the pre-test. However, the treatment group showed significant improvement on the post-test, which is apparently related to the

treatment they received after the pre-test.

5.2 Implications of the Study

Grammar learning is regarded as essential in foreign language acquisition (Purpura, 2004) and measuring language proficiency is crucial because it is regarded as the criteria for assessing learners' language skills (Alderson, 2005). With the advent of computer technology, immediate feedback can be delivered by a computer system, taking the test beyond the purpose of measuring students' ability to exploiting the test's potential as a learning tool.

As mentioned in previous chapters, the education system in Taiwan is more exam-oriented than in western educational settings. Thus, if we can understand what types of feedback on grammar tests are the most useful, there is a possibility students could achieve facilitative learning from the test.

Even though research on different types of feedback on grammar and other types of language tests is plentiful, little research has been done in Asia. Therefore, the findings of this study contribute to the field's understanding of different types of feedback on English grammar tests in Asian educational settings.

The results revealed that no significant difference was found between the mean scores of the pre-test and post-test within the control group. This finding suggests that the

teacher should give students corrective feedback because providing students with their scores without corrective feedback cannot help them improve their grammar proficiency.

The current study also found that there were significant differences in the pre-test and post-test mean scores in the treatment group, which received metalinguistic feedback. In other words, explicit corrective feedback, especially metalinguistic feedback, is likely to improve student performance on grammar proficiency tests. This finding indicates that metalinguistic feedback is useful and can be used by teachers to help students enhance their grammar proficiency.

In addition, because the feedback was immediately delivered by the test system, and the effect appears to have lasted for at least two weeks, this present study offers empirical evidence that metalinguistic explanations of grammatical rules are useful and computer technology makes the delivery of immediate feedback easier than in traditional testing. Hence, if instructors could make good use of technology and apply metalinguistic feedback in grammar exams, student learning should be facilitated.

5.3 Limitations of the Study and Suggestions for Future Research

The present study not only confirms the findings of previous studies with regard to types of grammar test feedback but also provides a better understanding of effectiveness of different types of grammar feedback in Asian second language learning settings.

However, there remain a number of limitations to the study, and these limitations may signal directions for future research.

One limitation is that this study only used quantitative data that focused on university freshmen's grammar performance on the pre-test and post-test. If qualitative data such as student perspectives on the feedback could be included, the researchers would have a more robust understanding of such feedback.

Another limitation is the lack of examination of student performance across different language proficiency levels. Though participants in the study came from three different levels of FENM classes, their data were only examined as two groups. Therefore, whether the improvement of student grammar proficiency differs because of different language levels could be explored in the future.

Finally, the number of participants in this study was limited. If future research on this topic could include more participants, the evidence would be strengthened. Further, the findings of the current study explored only a few aspects of the effectiveness of different types of feedback. Additional research is needed to ascertain other aspects of feedback.

REFERENCES

Abeywickrama, P. (2007). Measuring the knowledge of textual cohesion and coherence in learners of English as a second language (ESL). Los Angeles: The University of California Press.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

Alderson, J.C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing, 22*,301-320.

Bachman, L., & Palmer, A.S. (1996). *Language testing in practice.* New York: Oxford University Press.

Bachman, L. (2004). *Statistical Analysis for Language Assessment*. Cambridge: Cambridge University Press.

Bailey, K. (1998). *Learning about language assessment: Dilemmas, decisions, and directions.* Cambridge, MA: Heinle and Heinle.

Bhattarai, A. (2000). Consciousness-raising and teaching of English grammar in the governmental institutions of Nepal. *Journal of NELTA, 5*(2), 14-16.

Bitchener, J. & Young, S. & Cameron, D. (2005). The effect of different types of

corrective feedback on ESL student writing. *Journal of Second Language Writing 14*, 191-205.

Bitchener, J. & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12* (3), 409-431.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing 17*, 102-118.

Black, P. & Wiliam, D. (1998a). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan.80* (2), 139-148.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*, 653-675.

Brown, H.D. (2010). *Language assessment: Principles and classroom practices.* White Plains, New York: Pearson Education.

Brown, H.D. & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). White Plains, New York: Pearson Education.

Carroll, S. (2006). Silence, awareness and SLA. In M. G. O'Brien , C. Shea & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 17-24). Somerville, MA: Cascadilla Proceedings Project.

Cazden, C. R. (1974). Play with language and metalinguistic awareness: One dimension

of language experience. *The Urban Review, 7*, 28-29.

Chapelle, C.A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics, 19,* 254-272. doi: 10.1017/S0267190599190135

Chapelle, C.A. & Douglas, D. (2006). *Assessing language through computer technology.* Cambridge: Cambridge University Press.

Chen, G. Y. (2005). *Effects of pen-based and web-based tests on English testing performances of junior high school students*. Unpublished master's thesis, National Kaohsiung Normal University, Kaohsiung, Taiwan.

Chen, C.Y., Liou, H.C. & Chang, J.S. (2006). FAST- An automatic generation system for grammar tests. Paper presented at the International Conference on English Instruction and Assessment. Chiayi, Taiwan.

Collentine, J. (1998). Cognitive principles and CALL grammar instruction: A mind centered, input approach. *CALICO Journal, 15*(1-3), 1-18.

Cotos, E. & Pender,N. (2008). Automated diagnostic writing tests: Why? How? In C.A. Chapelle, Y. -R.Chung, &J. Xu (Eds.), *Toward adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 65-81).

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* (pp. 42-63). Cambridge:

Cambridge University Press.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 313-348). Malden, MA: Blackwell Publishing.

Doughty, C. (2003b). Effects of instruction on learning a second language: A critique of instructed SLA research. In B. VanPatten, J. Williams, & S. Rott (Eds.), *Form-Meaning Connections in Second Language Acquisition* (pp. 181-202). Mahwah, NJ: Lawrence Erlbaum.

Ellis, N. (2002). Frequency effects in language processing: A review with implication for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit knowledge. *Studies in Second Language Acquisition, 27*, 305-352.

Ellis, N. (2008). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education (*2nd ed., pp.1-13*).*

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

Ellis, R. & Loewen, S. & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28*, 339-368.

Ellis, R. (2008). Explicit form-focused instruction and second language acquisition. In B. Spolsky & F.M. Hult (Eds.), *The handbook of educational linguistics* (pp.437-455).

Fotos, S. (1993). Consciousness-raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics, 14*, 385-407.

Heift, T. (2001). Error-specific and individualized feedback in a web-based language tutoring system: Do they read it? *ReCALL, 13* (1), 99-109.

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL, 16* (2), 416-431.

"Item Analysis". *Classroom Assessment.* Retrieved June 6, 2012, from http://fcit.usf.edu/assessment/selected/responsec.html

Izumi, S. (2002). Output, input enhancement and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition, 24*, 541-577.

Jang, E.E. (2008). A framework for cognitive diagnostic assessment. In C.A. Chapelle, Y.-R.Chung, &J. Xu (Eds.), *Toward adaptive CALL: Natural language processing*

for diagnostic language assessment (pp. 117-131).

Jang, E.E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity argument for Fusion Model application to LanguEdge assessment. *Language Testing, 26* (1), 31-73.

Kitao, S.K. & Kiato, K. (1996). Testing grammar. *The Internet TESL Journal, 2*(6). Retrieved March 16, 2011, from http://iteslj.org/Articles/Kitao-TestingGrammar.html

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47*, 467 – 505.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition, 22*, 557 – 584.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, *27*(3), 405-430. doi:10.1093/applin/ami051

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *Modern Language Journal, 77*(3), 330-339.

Nagata, N., & Swisher, M. V. (1995). A study of consciousness-raising by computer: The effect of metalinguistic feedback on second language learning. *Foreign Language Annals, 28*(3), 337-347.

Nagata, N. (1995). An effective application of natural language processing in second

language instruction. *CALICO Journal, 13* (1), 47-67.

Nagata, N. (1997). The effectiveness of computer-assisted metalinguistic instruction: A case study in Japanese. *Foreign Language Annals, 30*(2), 187-200.

Nutta, J. (1998). Is computer-based grammar instruction as effective as teacher-directed grammar instruction for teaching L2 structures? *CALICO Journal, 16* (1), 49-62.

Purpura, J. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.

Rutherford, W.E. (1987). *Second language grammar: Learning and teaching.* New York: Longman.

Sharwood-Smith, M. (1981). Consciousness-raising and second language learner. *Applied Linguistics*, *2*, 159-169.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on attention and awareness in learning. In R. Schmidt, Ed., *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu, HI: University of Hawai`i, National Foreign Language Resource Center.

Sheen, R. (1992). Problem solving brought to task. *RELC Journal*, 23, 44-59.

Sims, J. (2006). The creation of a valid and reliable university proficiency exam. *Tunghai Journal of Humanities, 47*, 323-342.

Sugiharto, S. (2006). Grammar consciousness-raising: Research, theory, and application. *Indonesian Journal of English Language Teaching, 2*(2), 140-148.

Sugiharto, S. 2006. Consciousness-raising and the acquisition of the simple present tense rule. Paper presented at the sixth Malaysia International Conference on English Language Teaching (MICELT), Equatorial Hotel, Melaka, Malaysia.

Tsai, S. (1998). *The effects of cooperative learning on teaching English as a foreign language to senior high school students.* Unpublished master's thesis, National Kaoshiung Normal University, Kaoshiung, Taiwan.

Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.) *Handbook of research in second language teaching and learning* (pp. 671-691). Mahwah, NJ: Lawrence Erlbaum Associates.

Yeh, Y.C. (2004). *Students' perceptions of cooperative learning methods in one senior high school EFL classroom in Taiwan.* Unpublished master's thesis, Ming Chuan University, Taoyuan, Taiwan.

Yin,M. & Sims,J. & Cothran, D. (in Press). Scratching where they itch: Evaluation of feedback on a diagnostic English Grammar Test for Taiwanese university students *Language Assessment Quarterly: An International Journal.*

Appendix A

Grammar test item statistics and ratings of their Chinese and English explanations

Item number	Item facility	Item discrimination	Average rating of Chinese feedback	SD	Average rating of English feedback	SD
1	0.69	0.22	4.15	0.82	3.46	1.04
2	0.18	0.30	4.03	0.98	3.62	1.16
3	0.78	0.35	4.04	0.96	3.60	1.13
4	0.71	0.39	4.12	0.88	3.63	1.15
5	0.69	0.39	4.07	0.97	3.65	1.14
6	0.78	0.39	4.00	0.95	3.59	1.07
7	0.59	0.26	4.22	0.96	3.71	1.15
8	0.57	0.17	4.22	0.94	3.79	1.19
9	0.56	0.17	4.21	0.94	3.72	1.18
10	0.74	0.09	4.10	0.95	3.65	1.09
11	0.76	0.26	4.12	0.87	3.66	1.16
12	0.93	0.13	4.17	0.80	3.65	1.03
13	0.63	-0.04 [a]	4.13	0.93	3.71	1.17
14	0.69	0.61	4.18	0.83	3.66	1.17
15	0.66	0.43	4.04	0.95	3.56	1.12
16	0.87	0.09	3.98	0.95	3.53	1.18
17	0.60	0.17	4.28	0.75	3.74	1.09
18	0.19	0.17	4.37	0.76	3.79	1.19
19	0.34	-0.13 [a]	4.13	0.91	3.62	1.09
20	0.31	0.13	4.18	0.81	3.75	1.08
21	0.54	0.35	4.06	0.96	3.69	1.16
22	0.46	0.13	4.07	0.85	3.60	1.16
23	0.90	0.13	4.09	0.83	3.62	1.17
24	0.40	0.43	3.99	0.87	3.59	1.17
25	0.62	0.52	3.93	0.92	3.59	1.11
26	0.53	0.22	4.07	0.89	3.74	1.18
28	0.82	0.48	3.93	0.98	3.51	1.14
29	0.62	-0.04 [a]	3.91	0.86	3.59	1.14
30	0.50	0.57	3.91	1.09	3.65	1.19
31	0.51	0.22	4.16	0.84	3.66	1.15
32	0.49	0.43	3.97	0.86	3.69	1.10
33	0.54	0.22	4.10	0.95	3.66	1.10
34	0.40	0.39	4.06	0.94	3.69	1.14
35	0.47	0.65	4.01	0.94	3.50	1.11
36	0.47	0.26	4.01	0.95	3.62	1.22
37	0.72	0.48	4.10	0.89	3.66	1.12
38	0.69	0.43	4.13	0.80	3.67	1.08
39	0.38	0.13	4.18	0.90	3.96	1.12

Item number	Item facility	Item discrimination	Average rating of Chinese feedback	SD	Average rating of English feedback	SD
41	0.21	0.09	4.24	0.92	3.79	1.19
42	0.32	0.48	4.26	0.80	3.75	1.18
43	0.46	0.22	4.28	0.88	3.76	1.21
44	0.84	0.09	4.16	0.94	3.72	1.16
45	0.47	0.22	4.04	1.05	3.63	1.17
46	0.68	0.04	3.93	1.01	3.69	1.10
48	0.35	0.35	4.22	0.83	3.76	1.11
49	0.49	0.30	4.31	0.80	3.84	1.06
50	0.60	0.22	4.16	0.90	3.78	1.14
52	0.71	0.30	4.12	0.88	3.74	1.09
53	0.49	0.61	4.13	0.87	3.76	1.17
54	0.60	0.52	4.23	0.86	3.74	1.15
55	0.40	0.48	4.13	0.98	3.69	1.21
56	0.50	0.48	4.06	1.05	3.81	1.16
57	0.62	0.43	4.17	0.90	3.65	1.20
58	0.75	0.39	4.21	0.79	3.78	1.15
59	0.21	0.09	4.24	0.75	3.71	1.06
60	0.26	0.17	4.12	0.98	3.74	1.15

[a] Items with negative item discrimination were checked for content validity; it was decided that these items were testing their respective structures so were left in the test for analysis purposes.

Appendix B

Item Frequency (%) *of Pilot I Questions*

	Pre-test					Post-test				
Item	Key	A	B	C	D	Key	A	B	C	D
1	D	0.0	35	0.0	65.0	B	33.3	65.0	1.7	0.0
2	A	75.0	20.0	3.3	1.7	C	5.0	3.3	91.7	0.0
3	C	5.0	11.7	71.7	11.7	A	48.3	25.0	8.3	18.3
4	D	6.7	16.7	35.0	41.7	C	13.3	13.3	60.0	13.3
5	A	46.7	0.0	38.3	15.0	A	51.7	6.7	38.3	3.3
6	C	20.0	1.7	78.3	0.0	C	10.0	1.7	85.0	3.3
7	B	0.0	70.0	3.3	26.7	C	5.0	11.7	53.3	30.0
8	B	8.3	70.0	8.3	13.3	A	66.7	6.7	13.3	13.3
9	D	18.3	21.7	10.0	50.0	D	40.0	5.0	20.0	35.0
10	D	18.3	25.0	25.0	31.7	B	25.0	35.0	21.7	18.3
11	B	5.0	55.0	18.3	21.7	C	33.3	15.0	50.0	1.7
12	A	40.0	11.7	26.7	21.7	D	1.7	26.7	26.7	45.0
13	C	6.7	3.3	71.7	18.3	A	43.3	23.3	8.3	25.0
14	A	48.3	18.3	25.0	8.3	C	6.7	10.0	66.7	16.7
15	A	48.3	25.0	0.0	26.7	A	53.3	41.7	1.7	3.3
16	B	5.0	41.7	18.3	35.0	B	10.0	56.7	6.7	26.7
17	B	1.7	58.3	20.0	20.0	B	20.0	61.7	6.7	11.7
18	C	13.3	8.3	58.3	20.0	D	10.0	1.7	43.3	45.0
19	B	13.3	73.3	6.7	6.7	C	6.7	15.0	71.7	6.7
20	C	26.7	25.0	38.3	10.0	D	48.3	10.0	11.7	30.0
21	B	5.0	28.3	63.3	3.3	A	21.7	20.0	55.0	3.3
22	A	45.0	35.0	20.0	0.0	B	1.7	56.7	23.3	16.7
23	D	31.7	13.3	6.7	46.7	B	30.0	40.0	10.0	20.0
24	B	23.3	60.0	10.0	6.7	B	21.7	71.7	3.3	6.7
25	C	3.3	23.3	50.0	23.3	C	6.7	21.7	55.0	16.7
26	B	21.7	55.0	21.7	1.7	B	38.3	51.7	6.7	3.3
27	C	20.0	15.0	35.0	30.0	B	6.7	88.3	5.0	0.0
28	A	28.3	55.0	10.0	6.7	A	46.7	40.0	8.3	5.0
29	C	26.7	11.7	55.0	6.7	B	45.0	41.7	8.3	5.0
30	A	68.3	13.3	15.0	3.3	A	86.7	6.7	6.7	0.0

Appendix C

Item Frequency (%) *of the 6 revised items from Pilot I*

	Pre-test					Post-test				
Item	Key	A	B	C	D	Key	A	B	C	Key
3	C	5	11.7	71.7	11.7	A	48.3	25	8.3	A
7	B	0	70	3.3	26.7	C	5	11.7	53.3	C
9	D	18.3	21.7	10	50	D	40	5	20	D
13	C	6.7	3.3	71.7	18.3	A	43.3	23.3	8.3	A
27	C	20	15	35	30	B	6.7	88.3	5	B
30	A	68.3	13.3	15	3.3	A	86.7	6.7	6.7	A

Item Frequency (%) *of the 6 revised items from Pilot II*

	Pre-test					Post-test				
Item	Key	A	B	C	D	Key	A	B	C	Key
3	C	41.9	12.9	35.5	9.7	A	22.6	3.2	51.6	22.6
7	B	9.7	38.7	9.7	41.9	C	12.9	16.1	48.4	22.6
9	D	9.7	16.1	16.1	58.1	D	12.9	9.7	12.9	64.5
13	C	16.1	6.5	48.4	29	A	41.9	29	9.7	19.4
27	C	12.9	16.1	22.6	48.4	B	19.4	58.1	16.1	6.5
30	A	41.9	29	12.9	16.1	A	61.3	29	3.2	6.5

Item Frequency (%) *of the 6 revised items from Pilot III*

	Pre-test					Post-test				
Item	Key	A	B	C	D	Key	A	B	C	Key
3	C	7.9	7.9	84.2	0	C	10.5	10.5	78.9	0
7	B	2.6	52.6	0	44.7	C	2.6	13.2	78.9	5.3
9	D	10.5	26.3	18.4	44.7	D	21.1	10.5	21.1	47.4
13	C	5.3	0	55.3	39.5	A	60.5	5.3	13.2	21.1
27	C	18.4	13.2	42.1	26.3	C	2.6	5.3	86.8	5.3
30	A	71.1	13.2	7.9	7.9	A	92.1	0	5.3	2.6

Appendix D
Pre-test Feedback

A. WORD ORDER IN STATEMENTS, QUESTIONS, AND EXCLAMATIONS

2. She asked me ____D____
   a . where I to buy my shoes.
   b . where did I buy my shoes.
   c . where bought I my shoes.
   d . where I bought my shoes.

Explanation: B and C are not correct because the above statement is not a question, so you should use the regular sentence order (subject + verb). Although the word "asked" seems to indicate a question, here it is used to report a question that another speaker has asked. A is not correct for two possible reasons. First, if the subject "she" of the first clause is also the subject of the second clause, it would be correct to say "She asked me where to buy shoes" (i.e., recommend a good store that she should go to). Second, if the subject of the second clause is "I", then there should be no "to" before the verb "buy"; then the correct sentence should be "She asked me where I buy my shoes." By the way, the difference between "where I buy my shoes" and "where I bought my shoes" is that the former refers to where I regularly go to buy my shoes, while the latter refers to where I went (in the past) to buy my shoes.
B 和 C 並不正確，因爲本句並非疑問句，故你應使用一般的句子的順序（主詞＋動詞）。雖然這個字『asked』 似乎指出本句爲疑問句，但在本句卻是報導另一個說話者已經問的問題。A 是不正確的，其原因有二。首先，如果第一個句子的主詞 she 也是第二個句子的主詞，正確的說法應爲 she asked me where to buy shoes.(推薦給她一間不錯的店)。第二、如果第二個句子的主詞是『I』，那麼在 buy 之前就不須加 to，因此正確的句子是 She asked me where I buy my shoes.另外，子句 where I buy my shoes 跟 where I bought my shoes 的不同爲，前者是指我常去買鞋子的地方。後這是指我這雙鞋是去哪買的（我已買了這雙鞋）。

3. What ____A____!
   a . a big boy your son has become
   b . has your son become a big boy
   c . your son a big boy has become

d．has a big boy has become your son

Explanation: B and C are not correct because the above statement is not a question, so you should use the regular sentence order (subject + verb) after the word “what”. The speaker is surprised how big the boy has become, so “a big boy” should follow the word “what”. This is why A is correct and not C.

本句並非疑問句，故在 what 之後的子句只能用正常詞序（主詞＋動詞），故 B 和 D 不正確。此句是感嘆句，what 之後需接名詞，若爲單數名詞需接冠詞，故 A 正確、C 不正確。

B. SUBJECT-VERB AGREEMENT

B. 主詞、動詞的一致性

4. The old man standing under the park trees ____C____ happy.
   a．do not look
   b．not looking happy
   c．does not look happy
   d．not look happy

Explanation: The noun that is closest to the verb of the sentence is not always the subject. In this sentence, “trees” is not the subject but rather “the old man”. The subject “man” is singular, so the verb “look” is also singular.

在句子中緊跟動詞的名詞並非一定是主詞，題目中的”trees”非主詞、”the old man”才是主詞。因主詞”man”是單數，故動詞”look”也需用單數。

5. It is hard to believe that no one among our club's members ____D____ to pay their fee.
   a．to want
   b．wanting
   c．want
   d．wants

Explanation: The subject of the sentence is “no one”, which is singular. Therefore, the verb “want” should also be singular.

子句中的主詞”no one”是單數，故動詞”want”必須加 s。

C. COORDINATED CLAUSES

C. 對等連接詞子句

6. She isn't a full-time nurse in this hospital ____A____ a volunteer who helps out once a week.
   a . but rather
   b . and
   c . instead
   d . but also

Explanation: The sentence pattern is "not" + *verb*…but rather", so only A is possible. C would be possible if the sentence were divided into two sentences, "She isn't a full-time nurse. Instead, she is…"

本句句型是" not + verb….. but rather"，所以只有 A 是正確的答案。如果本句分成兩個句子，C 也可以選，如"She isn't a full-time nurse. Instead, she is….."

7. A man dressed in old, dirty clothes came to our door ____C____ for food.
   a . and begging
   b . but begging
   c . and begged
   d . but begged

Explanation: The coordinating conjunction "but" introduces a contrast or an unexpected outcome. If you say, "The man stopped but didn't say a word", the word "but" indicates that the second sentence is unexpected and surprises you. In the sentence above, the man's begging for food does not surprise you, so B and D are not correct. Secondly, when you use a coordinating conjunction like "and", the words in front and behind "and" should have the same form. In other words, if you put an adjective <u>before</u> "and" you should also put an adjective <u>after</u> "and". The verb "came" is in the past tense while "begging" is a present particle of a verb. They are not the same form, so A is not correct. C is correct because the verb "begged" is the same form as "came".

對等連接詞 but 表示一種相反或者預期之外的結果。假如你說：" The man stopped but didn't say a word"，那麼 but 這個字暗示" didn't say a word"使你驚訝以及在你的預期之外。題目中" the man's begging for food" 並沒有使你感到驚訝，所以 B、D 都不正

確。其次，當你使用對等連接詞時，句子兩端的結構要一致。換句話說，假如你在 and 之前使用形容詞，那麼也應該在 and 之後使用形容詞。本題在對等連接詞 and 之前使用過去式 came，故之後也應該使用過去式 begged，故選 C。

D. ADVERBIALS

D. 副詞

8. ____B____ the plane landed at the airport, we called home to say that we had arrived safely in Taiwan.
   a．Although
   b．As soon as
   c．Before
   d．While

Explanation: A is not possible since the word "although" introduces a contrast or an unexpected outcome. Answer A would be possible if the sentences were, "Although the plane landed at the airport, we were not allowed to disembark." The second sentence describes a situation we would not expect. C also does not make sense since you won't call to say that you have arrived until you have landed. D is not correct are not correct because "while" is normally followed by the continuous tense (i.e., we were getting). "As soon as" means "when" and is, therefore, the only answer that is logical.

A 不能選，因 although 表示一種相反或預期之外的結果，如果要用 although，則句子就要改寫成" Although the plane landed at the airport, we were not allowed to disembark."來表示一種預期之外的情形"。C 選項也不可能，因你無法在飛機降落之前，就打電話說你已經平安降落在台灣。 D 也不正確，因 while 通常用於進行時態之中（如 we were getting）。As soon as 的意思等於 when，故只能選 B。

9. ____B____ her last exam, Tina celebrated by going out to a movie with some friends.
   a．As she was taking
   b．Having taken
   c．Taking
   d．When she took

Explanation: Both A, C and D mean that she took the exam and celebrated at the same time. This is impossible, so those answers are not correct. B means "After she took" and is the only possible answer.

A、C 和 D 選項的意思是「考試」和「出外慶祝」同時發生，這是不可能的事，所以都不能選。B 選項的意思是" After she took"，這是唯一可能的答案。

E. CONDITIONALS

E. 條件句

10. ____D____, _____ at least ten kinds of nesting birds.
    a . If you walked through those woods now, you would have seen
    b . If you walk through those woods now, you would see
    c . If you walked through those woods now, you will see
    d . If you walk through those woods now, you will see

Explanation: In English, conditionals usually begin with the word "if" and can be labeled either *real conditions* and *unreal conditions*. Real conditions are those that can really happen while unreal conditions cannot. In real conditions, the first clause is in the present tense, and the second clause is in the future tense. In unreal conditions in the present, the first clause is in the past and the second clause uses "would". D is a real condition and is the only answer that is grammatically correct. A is an unreal condition but it is not grammatically correct. It would be correct if the answer were, "if you walked…, you would see".

在英文裡條件句通常由 if 來引導，可表達真實的情況或者是與事實相反的情況。真實的情況有可能發生，而與事實相反的情況則不可能發生。在表達真實的情況，句子前半段需用現在式，後半段需用未來式；而表達與現在事實相反，則前一個句子要用過去式，後一個句子要使用 would。故只有 D 符合句意和文法要求。A 是表達與現在事實相反的情況，但犯了文法的錯誤，正確的表達方式是：" if you walked……, you would see"。

11. At this moment Lisa doesn't remember all of the vocabulary on the quiz. ____D____.
    a . If she studied harder last night, she would remember it better
    b . If she had studied harder last night, she would have remembered it better
    c . If she studied harder last night, she would have remembered it better
    d . If she had studied harder last night, she would remember it better

Explanation: First, **t**his sentence is an *unreal condition* since she didn't study hard for the quiz and it is too late to change that now. Secondly, this sentence contains two different time frames. In the first clause, the time frame is in the past ("last night") ; in the second clause the time frame is in the present (at this moment when she is taking the quiz). Therefore, the first part is a past unreal condition ("had studied") and the second part is a present unreal condition ("would remember").

首先，本句爲『與事實相反』因爲她並沒有用功讀書準備考試，而且也不可能改變事實。第二、本句包括兩種時間條件。在第一個子句中，時間點爲『過去』last night。在第二個子句中，時間點爲現在-她考試的當下。因此，第一個子句爲與過去事實相反 (had studied)，第二個子句爲與現在事實相反 (would remember)。

F. RELATIVE CLAUSES

關係子句

12. John, ____B____ , was proud when he won first prize.
    a . whom we convinced him to join the race
    b . whom we convinced to join the race
    c . who we convinced him to join the race
    d . who we convinced to join the race

Explanation: It is important to notice two things here. First, the relative pronoun "who" is an object (i.e., we convinced "who"), so you should use "whom" in formal, written English. Secondly, the relative pronoun "whom" takes the place of John and, therefore, you should not add another reference to John by saying "him".

說明： 這裏有二個重點要注意。首先，此處的關係代名詞當受詞使用，因此，在正式英文書寫中，必需使用正確的受格格式"whom"。第二，由於此處的關代 "whom" 已取代 John 成爲其受格，故不可再填入另一個和"John"有關聯的代名詞 "him".

13. The Christmas decorations, ____A____, represented different scenes from the Biblical story of Jesus' birth.
    a . many of which Sylvia bought at the flea market
    b . many which Sylvia bought at the flea market
    c . which of many Sylvia bought at the flea market
    d . which many of them Sylvia bought at the flea market

Explanation: When you use a quantifier such as "many, much, some…" in a relative clause, you should use the structure "quantifier + of + relative pronoun" (i.e., many of which). This is the reason that only A is correct.
當使用量詞 (如 many, much, some…")在關係子句中時，必需遵守以下結構： 量詞 + of + 關代 (如： many of which). 故只有 A 是正確用法。

G. NOUN CLAUSES
G. 名詞子句

14. What advice can you give me about ____C____ a better job on my homework?
    a . that how I can
    b . how doing
    c . how to do
    d . how can I do

Explanation: "About" is a preposition, and only nouns follow prepositions, sentences such as questions. For this reason, D is not correct. D is a question. A is not correct because of the word "that. It would be correct if the answer were "how I can". The reduced form of "how I can" is "how to do", so C is "correct.
Explanation: sentences such as questions → sentences such as questions are not proper here.
" About" 在此當介係詞用，而其後只能放名詞類型而非問句類型。由於 D 是一個問句型態，故不正確。A 亦不正確是因爲多了"that"; 若把"that"去掉留下 " how I can", 那麼此句即正確。而選項 C 的 " how to do" 是 " how I can" 的簡化型態，故 C 爲正確。

15. ____A____ was a mystery to all of us.
    a . Why she was acting so strangely
    b . Why was she acting so strangely
    c . That why she was acting so strangely
    d . Why acting so strangely

Explanation: The answer to this one should be a noun because of the verb "was". You could be sure of this by putting a noun in the blank such as "her behavior" and, it makes sense. A noun clause is, therefore, appropriate. The structure of a noun clause is generally an *interrogative pronoun* (e.g., why) or *"*that*"* + *subject* + *verb*. Notice that you should

not put both the interrogative pronoun and “that” together in the same clause!
由於動詞是 ”was”的緣故，本題的答案必定爲名詞類。如要証實這個說法，你只要將一個如” her behavior”這樣的一個名詞置入空白處，即可得到一個有意義的正確句子。此外，放入一個名詞子句也同樣成立。名詞子句的結構，普遍來說是一個疑問代名詞或一個 “that” 加上 主詞 + 動詞。必需注意的是，疑問代名詞及 that 不能同時出現在同一個子句。

H. PAST TENSE
過去式

16. Charlie discovered that he ____A____ the wrong girl in the dark when the lights went on again.
    a . had kissed
    b . kissed
    c . is kissing
    d . was kissed

Explanation: The past perfect (“had kissed”) is used when there are two actions (verbs) in one sentence and the second action happens before the first. In this sentence Charlie kissed the wrong girl before he discovered this, so you should use the past perfect “had kissed”.
Explanation: in one sentence and …. → in one sentence (all happened in the past) and …
當二個動作(二個動詞)出現在同一個句子中(動作皆發生在過去)，而且第二個動作發生在第一個動作之前時，使用過去完成式 (had kissed)。在此句中，”Charlie 親錯女孩”是發生在他發現之前，故應該使用過去完成式(had kissed)。

17. I ____B ____ suspicious of Henry ever since I saw him come home late at night several days.
    a . am
    b . have been
    c . was
    d . had been

Explanation: This sentence is in the past, so A is not possible. Because of the word “since”, the first part of the sentence must be in the present perfect (“have been”). The present perfect expresses that the speaker has been suspicious from past until present. D

would be correct if the verb in the second part were “had seen”.
Explanation:
因爲本句是發生在過去，故 A 錯。由於“Since”的原因，此句的第一部份必需爲現在完成式 (have been)。現在完成式表示說話者從以前就一直對 Henry 起了疑心，直到現在也還是。在 D 句中，如果將第二部份的動詞改爲 “ had seen”，那麼 D 就正確。

I.PRESENT TENSE
現在式

18. Tom ____B ____ not to recognize which of the babies _____ to him.
    a . is appearing, is belonging
    b . appears, belongs
    c . appears, is belonging
    d . is appearing, belongs

Explanation: The verbs “appear” and “belong” cannot be used in a continuous tense, so only B is possible.
動詞 “appear” 和 “belong” 不能用在進行式的句型中，因此只有 B 是可能的。

19. Yum. I ____C ____ something good cooking in the kitchen. I wonder what it is.
    a . am smelling
    b . have been smelling
    c . smell
    d . smelled

Explanation: From the second sentence, it is clear that the person is speaking in the present, so D is not possible. The verb “smell” can be used in the present continuous and the simple present. When the verb “smell” is used in a continuous tense (e.g., I am smelling the flower.), it means that the speaker is inhaling close to an object in order to know its odor. This is <u>not</u> what the speaker means in this sentence, so A and B are not correct. When “smell” is used in the simple present or past, it means that the speaker is not making a conscious effort to perceive an odor. (S)he notices an odor suddenly and naturally. This is what the speaker means in this sentence, so D is correct.
Explanation: so D is correct. → so C is correct.
第二個句子裏清楚地表示出說話者的敘述是發生在現在，故選項 D 不可能。動詞 “smell” 可以用在現在進行式及現在簡單式。當 “smell”用在進行式(如 I am smelling the flower.)時，表示說話者正靠近地聞著某物件以期知道其氣味。然而這並不是本

題說話者的用意；因此 A 和 B 都不正確。而當 ”smell” 用在現在或過去簡單式時，表示說話者並非經由有意識地作爲而聞到氣味；而是突然、自然地注意到這個味道。這才是本句說話者的本意，故選 C。

J. FUTURE TENSE
未來式

20. Sally ____B ____ her grandfather tomorrow once classes _____ over.
    a . will visit, will be
    b . will visit, are
    c . visit, will be
    d . visit, are

Explanation: When you use the connector “once” to refer to the future, the verb before the connector should be in the future tense and then verb after the connector should be in the simple present.
當有連接詞 “once”的句型涉及到未來時態時，此連接詞之前的動詞應以未來式表示，而在此連接詞後的動詞則以簡單現在表示。

21. This time next week we ____C ____ on a beach enjoying the warm sunshine.
    a . will lie
    b . are going to lie
    c . will be lying
    d . will have lied

Explanation: The expression “this time” + *a time* requires the future continuous “will be lying”.
要表達未來之”此時”(如明天此時，下周此時…)，必需用未來進行式 “will be lying”。

K. MODALS
K. 情狀助動詞

22. Lucy ____B ____ to the party because she wasn’t in her room a few minutes ago.
    a . could go
    b . could have gone
    c . might go
    d . had to go

Explanation: This tense of the verbs in this sentence are clearly in the past. Lucy was not in her room because she probably went to the party. A is not correct because the tense is wrong. "Could" + *verb* is used to mean strong possibility in the *present and future*. "She could go" means that it is very possible that she will go. C is not correct because for the same reason. "Might" + *verb* is used to mean weak possibility in the *present and future*. D is the past tense of "must", but using "must" here doesn't make sense. B is correct because "could have gone to the party," means that the speaker is using logic to guess that she went to party. In other words, Lucy is not in her room, so I conclude that she went to the party. "Could + have" + *past participle* is used to mean strong possibility in the *past*.
解釋：這個句子的動詞時態很清楚的是過去式。Lucy 剛剛不在她的房間因爲她可能去了派對。A 是不正確的，因爲時態是錯的。"could"+ 動詞 是用來表示對現在和未來一種強烈的可能性。"She could go,"意思爲她非常可能將會去。C 是不正確的，理由與 A 相同。"might"+動詞 是用來表示對現在和未來一種較小的可能性。D 是"must"的過去時態，但"must"在這不符合句意。B 是正確的，因爲"could have gone to the party,"意思爲說話的人用邏輯的方式去臆測她去了派對。換句話說，Lucy 現在不在她的房間，所以我斷定她去了派對。"could have"+ 過去分詞 是用來表示對過去事情強烈的可能性。

23. After having asked the coach for many weeks, I ____A ____ the team yesterday.
    a . was finally able to join
    b . could finally join
    c . could have finally joined
    d . can finally join

Explanation: The word "yesterday" indicates that the sentence is in the past, so D is not correct. C is not correct because "could have joined," means that the speaker is using logic to guess something. For example, "she could have gotten lost because she hasn't arrived yet" means that you conclude logically that she is lost because she hasn't arrived at her destination. However, in the above sentence the speaker is saying this about himself, which would be strange. "B is not correct because "could" is used to mean ability in the past during a period of time. For example, "I could jump very high when I was young." "Could" is not used for a specific time in the past. You cannot say, for example, "I could jump high yesterday." In this case, you need to say, "I was able to jump high yesterday." That is the reason A is correct.
解釋："yesterday"這個字指出這個句子的時態爲過去式，所以 D 是不正確的。C 是不正確的，因爲"could have joined"意思爲說話的人用邏輯的方式去臆測某事。例如，

“She could have gotten lost because she hasn’t arrived yet,”意思爲你邏輯地斷定她迷路了因爲她還沒到達她的目的地。然而，上述句子若說話的人用來臆測自己的事會非常奇怪。B 是不正確的，因爲“could”是用來表示在過去一段時間時的能力。例如，“I could jump very high when I was young.”“could”不是用來表示過去一特定的時間。你不能說“I could jump high yesterday.”，而是要說“I was able to jump high yesterday.”。這就是爲什麼 A 是正確的。

L. COMPARISON

L. 比較

24. Having a pool party is ____D ____ having it at the beach!
    a . not the same fun as
    b . quite less fun as
    c . as not fun as
    d . not quite as fun as

Explanation: Only D is grammatically correct. B would be correct if the answer were “less fun than”. C would be correct if the answer were “not as fun as”.

解釋：只有 D 是合乎文法且正確的。B 的答案需改爲“less fun than”。C 的答案需改爲“not as fun as”。

25. ____B ____ than any other job I know.
    a . An international tour guide is more stressful
    b . Being an international tour guide is more stressful
    c . An international tour guide is stressfuller
    d . Being an international tour guide is more stressfulller

Explanation: Only a job can be stressful, not a person, so A and C are not correct. “Being an international tour guide” means *working as an international tour guide*. The comparative form of “stress” is *more stressful*, so B is correct.

解釋：stressful (有壓力的) 這個字的主詞應該是工作而不是人，所以 A 和 C 都不正確。“Being an international tour guide”意思爲*以國際導遊爲職*。“stress”的比較級爲 *more stressful*，所以 B 是正確的。

M. ARTICLES

M. 冠詞

26. She teaches in ____C ____ Biology department of _____Tunghai University.
    a . the, the
    b . (nothing), (nothing)
    c . the, (nothing)
    d . (nothing), the

Explanation: In front of college departments (…Biology department), you should use "the", but not in front of the names of universities (Tunghai University). There is, however, an exception. When a university is followed by "of ", you need to add the article "the" such as in the names "the University of California" and "the State College of Michigan".
解釋：在大學科系前需使用定冠詞"the"（…Biology department 生物系），但是在大學校名前則不用（Tunghai University 東海大學）。然而例外的情形爲，當大學校名有"of"則須加用定冠詞"the"，例如，"the University of California"和"the State College of Michigan"。

27. We were very unhappy because ____B ____ suitcases that we had put in the bus's storage compartment were scratched.
    a . some of
    b . some of the
    c . some
    d . (nothing)

Explanation: A, C and D are not correct because of the relative clause "that we had put…". The relative clause makes it clear which suitcases the speaker means and, therefore, you must add "the" to show this. The article "the" is used to indicate that the suitcases are specific suitcases and not just any suitcases.
解釋：A，C 和 D 都不正確。因爲關係子句"that we had put…"很清楚告訴我們說話者所指的行李箱爲何，因此必須使用定冠詞"the"。冠詞"the"是用來指出這些行李箱是特定的行李箱，而不是任何一些行李箱。

N. PASSIVES
N. 被動式

28. During the earthquake, Mr. Peterson ____C ____ when a bookcase fell down on him.
    a . hurt

b．hurted
c．was hurt
d．was hurted

Explanation: In this sentence the subject “Mr. Peterson” did not hurt himself; the falling bookcase hurt him, so the idea expressed here is that “Mr. Peterson was hurt by the falling bookcase.” This sentence should, therefore, be in the passive voice. The past participle of the verb “hurt” is also “hurt”.
解釋：這個句子的主詞“Mr. Perterson”並沒有傷他自己，而是倒下來的書櫥傷了他，所以這裡的句意爲“Mr. Perterson was hurt by the falling bookcase.”Mr. Perterson 被倒下來的書櫥傷了他。因此這個句子須用被動語態。動詞“hurt”的過去分詞也是“hurt”。

29. The army general was afraid that some of his soldiers ____A ____ during the battle.
a．might have been captured
b．might be captured
c．might have captured
d．might capture

Explanation: In this sentence the soldiers didn’t capture themselves; someone captured them. Those people are not mentioned, but we can guess that they must be “the enemy”. Therefore, the sentence means, “the soldiers were captured by the enemy” and must be in the passive voice. When you use a modal such as “may or might” in the passive voice in the past tense, the verb after the modal changes to “have been…”.
解釋：在這個句子中，士兵們不會捕捉他們自己，而是某人捕捉他們。某人是誰，句中並沒提及，但是我們可以猜他們必定爲敵人。因此這裡的句意爲“the soldiers were captured by the enemy”士兵被敵人捕捉，而且必定爲被動語態。當我們使用情狀助動詞“may 或 might”在被動語態的過去式時，被動語態之後的動詞須改爲“have been…”。

O. INFINITIVES AND GERUNDS
O. 不定詞和動名詞

30. The neighbors don’t remember ___C ____ anyone leave the house the night the murder took place.
a．to see
b．to have seen
c．seeing
d．have seen

Explanation: There are only three things that you can put after the verb "remember": 1) "that" + *a sentence*, 2) *an infinitive* or 3) *a gerund*. For this reason, D is not correct. When a speaker uses an infinitive after "remember", he means that he didn't forget to do something. For example, someone has expressed doubt that you locked the door before leaving the house, so you reassure him by saying, "Don't worry. I remembered <u>to lock</u> the door before leaving the house." The above sentence does not express this, so, A and B are not correct. When a speaker uses a gerund after "remember", he means that he can still remember an event, situation or action. For example, I remember arriving home at 5:00." Here the speaker means that he still remembers the action of arriving home at 5:00. Similarly, in the sentence above, the speaker means that the neighbors don't remember the action of someone leaving the house. For this reason C is the correct.
解釋：動詞"remember"之後只可接三種情況：1)"that"+子句， 2) 不定詞，3) 動名詞。因此 D 是不正確的。當說話者在"remember"之後使用不定詞，意思爲他沒有忘記去做某事。例如，有人懷疑你出門前是否鎖了門因此你可以跟他保證說，"Don't worry. I remembered <u>to lock</u> the door before leaving the house."不要擔心，我記得我出門前鎖了門了。上述句子所表達的並非如此，因此 A 和 B 都不正確。當說話者在"remember"之後使用動名詞，意思爲他仍然記得一件事情，情況或舉動。例如，"I remember arriving home at 5:00."我記得我五點到家。這裡說話者的意思爲，他仍然記得他到家這個舉動是在五點鐘。相似地，上述的句子中，說話者的意思爲，鄰居們不記得有人離開屋子這個舉動。因此 C 是正確的。

31. ____A ____ left at that last intersection was a big mistake
    a . Turning
    b . He turned
    c . Turned
    d . He turns

Explanation: B and D are incorrect because there is already a verb in this sentence – "was". To understand this better, we could simplify the sentence by changing it into, "<u>*This action*</u> was a big mistake." Therefore, the part of the sentence that goes in the blank should be a noun. To change the verb "turn" into a noun, you can add "ing" to it so that it becomes the gerund "turning". This is why A is the only correct answer.
解釋：B 和 D 都不正確因爲這個句子已經有動詞"was"。爲了更了解這個句子，我們可以將它簡化改變爲"<u>*This action*</u> was a big mistake."。因此，句中的空格應填入名詞。爲了將動詞"turn"改爲名詞，可加上"ing"改爲動名詞"turning"。這就是爲什麼 A 是唯一的正確答案。

Appendix E

Instruction Sheet for Participants (High-level class)—Chinese Version

各位同學你們好：
首先，非常感謝你們協助完成此項線上文法測驗。此項測驗爲協助了解不同詳細度的解答對文法能力進步的影響。測驗總共分成兩次，每次內容爲三十題文法選擇題。每次作答時間爲三十分鐘。待測驗完畢後將會提供每位同學一份小禮物以致謝意。本次測驗的數據將作爲研究用途，不會影響同學的學期成績，請同學放心作答。

<u>爲了讓同學事先了解測驗的流程及須注意的事項，以下說明請同學仔細閱讀。</u>

(1) 測驗將分成兩個部分，分別在 10/3（一）及 10/17（一）早上第一二節進行。
(2) 爲達測驗的完整性，請同學這兩天都務必出席。
(3) 因在測驗前後皆需要時間進行說明，請同學務必準時到指定測驗地點。
(4) 因需使用登入東海師生資訊系統的帳號密碼登入測驗，請同學事先確認。
(5) 每一班同學將會被分成兩個組別，請按照分組名單及指定時間到電腦教室（M023）進行測驗。非測驗時段同學請在原教室聽從老師安排。
(6) 每位同學所屬的組別及測驗的時間不同，請務必在名單上找到自己的學號，並依照指示進行。
(7) 第一組同學進行完測驗後即可離開。第二組同學測驗後會需要留下來閱讀詳解，並在記錄表上記下所花時間。（測驗當天會詳細說明）

分組名單：

Group 1	**8:10~9:00**	**Group 2**	**9:10~10:00**
01004015	01004049	01008014	01008046
01004606	01004005	01004310	01004322
01004923	01004059	01004026	01004601
01008067	01004309	01004902	01004616
01004642	01008061	01008011	01008060
01004018	01004024	01008054	01004943
01004032	01004306	01008004	01004067
01004316	01004608	01004064	

Appendix F

Instruction Sheet for Participants (High-level class)—English Version

Dear all,

First of all, thanks for your help with the online grammar test. This test Helps the researcher know the effectiveness of different types of feedback on the online grammar test. The test is divided into two parts, and is administered on two different days. In each part, there are 30 grammar multiple choice questions, and you will have 30 minutes to answer the questions. After the test, each of you will receive a gift to represent the researcher's thanks. The data of this test will be used only for research purpose, and will not affect your grade.

◆ **Before the test, please read the following instructions carefully.**

(1) The test is divided into two parts, and is administered on 10/3 and 10/17 Mondays from 8:10 to 10:00.

(2) To help get a complete data, please participate in both of the two tests.

(3) Please arrive at the computer room (M023) on time because we need time to give instructions.

(4) Please check with your Tunghai account number and password in advance, we will need them to log in the test system.

(5) Each class will be divided into two groups. You can find which group you are assigned to from the chart below.

(6) After you know which group you are in, please follow the instructions. If it is not your turn to take the test, please stay in class.

(7) For students in Group 1, you may leave after finishing the questions.

For students in Group 2, you will need to stay after finishing the questions, and read the feedback on the system. Also, you will need to record how much time you spend on reading the feedback. (More information will be given on the test day)

Group List：

Group 1	**8:10~9:00**	**Group 2**	**9:10~10:00**
01004015	01004049	01008014	01008046
01004606	01004005	01004310	01004322
01004923	01004059	01004026	01004601
01008067	01004309	01004902	01004616
01004642	01008061	01008011	01008060
01004018	01004024	01008054	01004943
01004032	01004306	01008004	01004067
01004316	01004608	01004064	

Appendix G

Instruction Sheet for Participants (Mid-level class)—Chinese Version

各位同學你們好：
首先，非常感謝你們協助完成此項線上文法測驗。此項測驗爲協助了解不同詳細度的解答對文法能力進步的影響。測驗總共分成兩次，每次內容爲三十題文法選擇題。每次作答時間爲三十分鐘。待測驗完畢後將會提供每位同學一份小禮物以致謝意。本次測驗的數據將作爲研究用途，不會影響同學的學期成績，請同學放心作答。

爲了讓同學事先了解測驗的流程及須注意的事項，以下說明請同學仔細閱讀。

(1) 測驗將分成兩個部分，分別在 10/3（一）及 10/17（一）早上第一二節進行。
(2) 爲達測驗的完整性，請同學這兩天都務必出席。
(3) 因在測驗前後皆需要時間進行說明，請同學務必準時到指定測驗地點。
(4) 因需使用登入東海師生資訊系統的帳號密碼登入測驗，請同學事先確認。
(5) 每一班同學將會被分成兩個組別，請按照分組名單及指定時間到電腦教室（M023）進行測驗。非測驗時段同學請在原教室聽從老師安排。
(6) 每位同學所屬的組別及測驗的時間不同，請務必在名單上找到自己的學號，並依照指示進行。
(7) 第一組同學進行完測驗後即可離開。第二組同學測驗後會需要留下來閱讀詳解，並在記錄表上記下所花時間。（測驗當天會詳細說明）

分組名單：

Group 1	8:10~9:00	Group 2	9:10~10:00
01004953	01004631	01008059	01004804
01008040	01008045	01004031	01004723
01004940	01004936	01008009	01008027
01004758	01004004	01004817	01004652
01004046	01004359	01008024	01004851
01004045	01004945	01008003	01004947
01004646	01004905	01004333	01004603
01004811		01004946	

Appendix H

Instruction Sheet for Participants (Mid-level class)—English Version

Dear all,

First of all, thanks for your help with the online grammar test. This test Helps the researcher know the effectiveness of different types of feedback on the online grammar test. The test is divided into two parts, and is administered on two different days. In each part, there are 30 grammar multiple choice questions, and you will have 30 minutes to answer the questions. After the test, each of you will receive a gift to represent the researcher's thanks. The data of this test will be used only for research purpose, and will not affect your grade.

◆ **Before the test, please read the following instructions carefully.**

(8) The test is divided into two parts, and is administered on 10/3 and 10/17 Mondays from 8:10 to 10:00.

(9) To help get a complete data, please participate in both of the two tests.

(10) Please arrive at the computer room (M023) on time because we need time to give instructions.

(11) Please check with your Tunghai account number and password in advance, we will need them to log in the test system.

(12) Each class will be divided into two groups. You can find which group you are assigned to from the chart below.

(13) After you know which group you are in, please follow the instructions. If it is not your turn to take the test, please stay in class.

(14) For students in Group 1, you may leave after finishing the questions.

For students in Group 2, you will need to stay after finishing the questions, and read the feedback on the system. Also, you will need to record how much time you spend on reading the feedback. (More information will be given on the test day)

Group List:

Group 1	8:10~9:00	Group 2	9:10~10:00
01004953	01004631	01008059	01004804
01008040	01008045	01004031	01004723
01004940	01004936	01008009	01008027
01004758	01004004	01004817	01004652
01004046	01004359	01008024	01004851
01004045	01004945	01008003	01004947
01004646	01004905	01004333	01004603
01004811		01004946	

Appendix I

Instruction Sheet for Participants (Low-level class)—Chinese Version

各位同學你們好：
首先，非常感謝你們協助完成此項線上文法測驗。此項測驗爲協助了解不同詳細度的解答對文法能力進步的影響。測驗總共分成兩次，每次內容爲三十題文法選擇題。每次作答時間爲三十分鐘。待測驗完畢後將會提供每位同學一份小禮物以致謝意。本次測驗的數據將作爲研究用途，不會影響同學的學期成績，請同學放心作答。

<u>爲了讓同學事先了解測驗的流程及須注意的事項，以下說明請同學仔細閱讀。</u>

(1) 測驗將分成兩個部分，分別在 10/3（一）及 10/17（一）早上第一二節進行。
(2) 爲達測驗的完整性，請同學這兩天都務必出席。
(3) 因在測驗前後皆需要時間進行說明，請同學務必準時到指定測驗地點。
(4) 因需使用登入東海師生資訊系統的帳號密碼登入測驗，請同學事先確認。
(5) 每一班同學將會被分成兩個組別，請按照分組名單及指定時間到電腦教室（M023）進行測驗。非測驗時段同學請在原教室聽從老師安排。
(6) 每位同學所屬的組別及測驗的時間不同，請務必在名單上找到自己的學號，並依照指示進行。
(7) 第一組同學進行完測驗後即可離開。第二組同學測驗後會需要留下來閱讀詳解，並在記錄表上記下所花時間。（測驗當天會詳細說明）

分組名單：

Group 1	8:10~9:00	Group 2	9:10~10:00
01004355	01008038	01004715	01004051
01004937	01008039	01004742	01004042
01004605	01004003	01004346	01004006
01004738	01004930	01004837	01004302
01004746	01004039	01004629	01004749
01004941	01004362	01004822	01004040
01004740		01004661	

Appendix J

Instruction Sheet for Participants (Low-level class)—English Version

Dear all,

First of all, thanks for your help with the online grammar test. This test Helps the researcher know the effectiveness of different types of feedback on the online grammar test. The test is divided into two parts, and is administered on two different days. In each part, there are 30 grammar multiple choice questions, and you will have 30 minutes to answer the questions. After the test, each of you will receive a gift to represent the researcher's thanks. The data of this test will be used only for research purpose, and will not affect your grade.

◆ **Before the test, please read the following instructions carefully.**

(15) The test is divided into two parts, and is administered on 10/3 and 10/17 Mondays from 8:10 to 10:00.

(16) To help get a complete data, please participate in both of the two tests.

(17) Please arrive at the computer room (M023) on time because we need time to give instructions.

(18) Please check with your Tunghai account number and password in advance, we will need them to log in the test system.

(19) Each class will be divided into two groups. You can find which group you are assigned to from the chart below.

(20) After you know which group you are in, please follow the instructions. If it is not your turn to take the test, please stay in class.

(21) For students in Group 1, you may leave after finishing the questions. For students in Group 2, you will need to stay after finishing the questions, and read the feedback on the system. Also, you will need to record how much time you spend on reading the feedback. (More information will be given on the test day)

Group List：

Group 1	**8:10~9:00**	**Group 2**	**9:10~10:00**
01004355	01008038	01004715	01004051
01004937	01008039	01004742	01004042
01004605	01004003	01004346	01004006
01004738	01004930	01004837	01004302
01004746	01004039	01004629	01004749
01004941	01004362	01004822	01004040
01004740		01004661	

Appendix K

Record Sheet for Participants (Treatment Group)

詳解閱讀時間記錄表

說明：請第二組受試同學在下面空格處依序填上：1) 基本資料 2) 開始及結束閱讀詳解的時間。

科系：_____

學號：_____

大一英文班別：_____

開始閱讀詳解時間：_____點_____分

結束閱讀詳解時間：_____點_____分

Appendix L

Pre-test and Post-test Questions

Item	Pre-test	Post-test
1.	She asked me _____ a. where I to buy my shoes. b. where did I buy my shoes. c. where bought I my shoes. d. where I bought my shoes.	He asked me _____ a. what time was it. b. what time it was. c. what it was time. d. what was it time.
2.	What _____! a. big boy your son has become b. has your son become a big boy c. your son a big boy has become d. has a big boy has become your son	What_____! a. have you become a beautiful woman b. you have become a beautiful woman c. a beautiful woman you have become d. have you a beautiful woman become
3.	The old man standing under the park trees _____ happy. a. do not look b. not looking happy c. does not look happy d. not look happy	The teacher talking to the students_____strict. a. do not look b. not looking c. does not look d. not look
4.	It is hard to believe that no one among our club's members _____ to pay their fee. a. to want b. wanting c. want d. wants	It seems like no one in our family _____the dress I just bought. a. to like b. liking c. likes d. like
5.	She isn't a full-time nurse in this hospital _____ a volunteer who helps out once a week. a. but rather b. and c. instead d. but also	He's not going to a movie with us this afternoon,_____, he'll stay at home and prepare for midterm. a. but rather b. and c. instead d. but also

6.	A man dressed in old, dirty clothes came to our door _____ for food. a. and begging b. but begging c. and begged d. but begged	The mysterious woman smiled, turned around, _____in the crowd. a. and disappearing b. but disappearing c. and disappeared d. but disappeared
7.	_____ the plane landed at the airport, we called home to say that we had arrived safely in Taiwan. a. Although b. As soon as c. Before d. While	Check it carefully_____you hand it in. a. although b. as soon as c. before d. after
8.	_____ her last exam, Tina celebrated by going out to a movie with some friends. a. As she was taking b. Having taken c. Taking d. When she took	_____ her children to school, the mother went back to sleep. a. Having driven b. As she was driving c. Driving d. when she drove
9.	_____, _____ at least ten kinds of nesting birds. a. If you walked through those woods now, you would have seen b. If you walk through those woods now, you would see c. If you walked through those woods now, you will see d. If you walk through those woods now, you will see	_____, _____to the park a. If the weather were nice today, I would go b. If the weather were nice today, I would went c. If the weather is nice today, I would go d. If the weather is nice today, I will go

10.	At this moment Lisa doesn't remember all of the vocabulary on the quiz. _____. a. If she studied harder last night, she would remember it better b. If she had studied harder last night, she would have remembered it better c. If she studied harder last night, she would have remembered it better d. If she had studied harder last night, she would remember it better	The view was wonderful._____. a. If I had a camera with me, I would have taken some photographs. b. If I had brought a camera with me, I would take some photographs. c. If I had brought a camera with me, I would have taken some photographs. d. If I had a camera with me, I would have taken some photographs.
11.	John, _____ , was proud when he won first prize. a. whom we convinced him to join the race b. whom we convinced to join the race c. who we convinced him to join the race d. who we convinced to join the race	Jenny, _____, was friendly when I asked her for help. a. who I met on campus yesterday b. who I met her on campus yesterday c. whom I met on campus yesterday d. whom I met her on campus yesterday
12.	The Christmas decorations, _____, represented different scenes from the Biblical story of Jesus' birth. a. many of which Sylvia bought at the flea market b. many which Sylvia bought at the flea market c. which of many Sylvia bought at the flea market d. which many of them Sylvia bought at the flea market	The paintings, _____, were symbols of his appreciation of art. a. many which they James bought b. which of many James bought c. which many of them James bought d. many of which James bought
13.	What advice can you give me about _____ a better job on my homework? a. that how I can b. how doing c. how to do d. how can I do	What suggestion can you give me about_____for your sister's birthday? a. what to buy b. that I buy c. what buying d. what I buy

14.	_____ was a mystery to all of us. a. Why she was acting so strangely b. Why was she acting so strangely c. That why she was acting so strangely d. Why acting so strangely	_____was hot gossip in our town a. Why married the crazy woman he b. Why marrying the crazy woman c. Why he married the crazy woman d. Why did he marry the crazy woman.
15.	Charlie discovered that he _____ the wrong girl in the dark when the lights went on again. a. had kissed b. kissed c. is kissing d. was kissed	Carrie found that she _____the wrong book in the bookstore when she returned home. a. had bought b. bought c. buys d. is buying
16.	I _____ suspicious of Henry ever since I saw him come home late at night several days. a. am b. have been c. was d. had been	Angie_____attracted to Michael from the first time they met each other to now. a. is b. has been c. has d. had been
17.	Tom _____ not to recognize which of the babies _____ to him. a. is appearing, is belonging b. appears, belongs c. appears, is belonging d. is appearing, belongs	The evidence ____to show that the responsibility for the mistake_____to him. a. appears, is belonging b. appears, belongs c. is appearing, is belonging d. is appearing, belongs
18.	Yum. I _____ something good cooking in the kitchen. I wonder what it is. a. am smelling b. have been smelling c. smell d. smelled	I _____something scary in the dark house. I wonder what it is. a. am hearing b. have been hearing c. heard d. hear

19.	Sally _____ her grandfather tomorrow once classes _____ over. a. will visit, will be b. will visit, are c. visit, will be d. visit, are	I _____you once I _____at school a. call, will arrive b. will call, arrived c. will call, arrive d. call, arrive
20.	This time next week we _____ on a beach enjoying the warm sunshine. a. will lie b. are going to lie c. will be lying d. will have lied	This time tomorrow I _____lunch with my friend Chris, having a nice chat. a. will eat b. are going to eat c. will have eaten d. will be eating
21.	Lucy _____ to the party because she wasn't in her room a few minutes ago. a. could go b. could have gone c. might go d. had to go	Jessica_____earlier because she told me she's not hungry a few minutes ago. a. could have had lunch b. could have lunch c. might have lunch d. has to have lunch
22.	After having asked the coach for many weeks, I _____ the team yesterday. a. was finally able to join b. could finally join c. could have finally joined d. can finally join	After saving money for a long time, I _____the camera yesterday. a. should finally buy b. was finally able to buy c. could have finally bought d. can finally buy
23.	Having a pool party is _____ having it at the beach! a. not the same fun as b. quite less fun as c. as not fun as d. not quite as fun as	Watching a movie at home is _____watching it in a movie theater. a. quite less exciting as b. not quite as exciting as c. as not exciting as d. not the same exciting as

24.	_____ than any other job I know. a. An international tour guide is more stressful b. Being an international tour guide is more stressful c. An international tour guide is stressfuller d. Being an international tour guide is more stressfulller	_____ than any other job I know. a. A doctor is more stressful b. Being a doctor is more stressful c. A doctor is stressfuller d. Being a doctor is more stressfuller
25.	She teaches in _____ Biology department of _____Tunghai University. a. the, the b. (nothing), (nothing) c. the, (nothing) d. (nothing), the	He studies in _____Social Work department of _____Taiwan University a. the, the b. (nothing), (nothing) c. the, (nothing) d. (nothing), the
26.	We were very unhappy because _____ suitcases that we had put in the bus's storage compartment were scratched. a. some of b. some of the c. some d. (nothing)	The teacher was very happy because_____students in his class went to good universities. a. most of b. most of the c. most d. (nothing)
27.	During the earthquake, Mr. Peterson _____ when a bookcase fell down on him. a. hurt b. hurted c. was hurt d. was hurted	The window of the building _____in a storm a few days ago. a. broke b. breaked c. was broken d. was breaked

28.	The army general was afraid that some of his soldiers _____ during the battle. a. might have been captured b. might be captured c. might have captured d. might capture	I haven't received the letter. It _______to the wrong address. a. might have been sent b. might be sent c. might have sent d. might sent
29.	The neighbors don't remember _____ anyone leave the house the night the murder took place. a. to see b. to have seen c. seeing d. have seen	You didn't remember _______the air conditioner before you left this morning. a. to have turned off b. turn off c. turning off d. to turn off
30.	_____ left at that last intersection was a big mistake a. Turning b. He turned c. Turned d. He turns	_____class is a bad behavior. a. Skipping b. He skip c. Skipped d. He skips

Appendix M

Participants' Total Score on the Pre-test and the Post-test

Group 1 (N=45)			Group 2 (N=45)		
ID	**Pre-test**	**Post-test**	**ID**	**Pre-test**	**Post-test**
1004003	18	19	1004006	15	21
1004004	15	10	1004020	24	25
1004005	23	18	1004026	21	23
1004015	20	19	1004031	16	18
1004018	16	19	1004040	24	20
1004024	21	18	1004041	16	24
1004032	21	23	1004042	11	16
1004039	8	12	1004051	17	20
1004046	18	16	1004064	13	22
1004049	24	20	1004067	14	19
1004059	18	17	1004302	18	16
1004306	22	23	1004310	21	23
1004309	22	20	1004322	22	24
1004316	21	22	1004333	18	26
1004355	8	13	1004346	25	24
1004359	20	19	1004601	16	21
1004362	7	12	1004603	17	22
1004605	17	14	1004616	18	22
1004606	19	22	1004629	16	19
1004608	17	14	1004652	19	20
1004631	20	16	1004661	13	20
1004642	20	22	1004715	15	20
1004646	14	18	1004723	18	18
1004738	13	16	1004742	10	20
1004740	16	17	1004749	12	17
1004746	17	21	1004804	15	17
1004758	20	16	1004822	16	12
1004811	13	16	1004837	16	17
1004905	19	18	1004851	12	12
1004923	20	21	1004902	19	28
1004930	12	16	1004943	23	25
1004936	17	16	1004946	18	17
1004937	23	20	1004947	15	18

Group 1 (N=45)	Group 2 (N=45)	Group 1 (N=45)	Group 2 (N=45)	Group 1 (N=45)	Group 2 (N=45)
1004940	23	19	1008004	17	24
1004941	15	17	1008007	21	24
1004945	18	21	1008009	15	19
1004953	12	19	1008011	18	21
1008038	17	16	1008014	24	19
1008039	15	14	1008024	16	22
1008040	17	20	1008027	18	22
1008043	21	21	1008046	23	21
1008045	21	22	1008054	21	25
1008055	15	25	1008059	19	21
1008061	13	15	1008060	19	21
1008067	23	26	991967	12	14

Appendix N

Item Frequency (%) of the Pre-test and the Post-test

	Pre-test					Post-test				
Item	Key	A	B	C	D	Key	A	B	C	D
1	D	0.0	24.0	0.0	66.0	B	15.0	71.0	3.0	1.0
2	A	71.0	16.0	2.0	1.0	C	2.0	2.0	86.0	0.0
3	C	4.0	12.0	70.0	4.0	C	8.0	16.0	59.0	7.0
4	D	2.0	7.0	29.0	52.0	C	11.0	8.0	55.0	16.0
5	A	35.0	3.0	46.0	6.0	A	44.0	8.0	32.0	
6	C	17.0	1.0	71.0	1.0	C	7.0	0.0	81.0	2.0
7	B	1.0	75.0	0.0	14.0	C	2.0	11.0	75.0	2.0
8	B	6.0	66.0	8.0	10.0	A	62.0	5.0	15.0	8.0
9	D	14.0	20.0	6.0	50.0	D	32.0	2.0	9.0	47.0
10	D	21.0	32.0	14.0	23.0	B	21.0	36.0	26.0	7.0
11	B	7.0	46.0	6.0	31.0	C	34.0	3.0	52.0	1.0
12	A	47.0	1.0	24.0	18.0	D	1.0	17.0	19.0	53.0
13	C	3.0	3.0	68.0	16.0	A	65.0	4.0	3.0	18.0
14	A	51.0	17.0	16.0	6.0	C	2.0	5.0	61.0	22.0
15	A	46.0	26.0	0.0	18.0	A	59.0	31.0	0.0	0.0
16	B	5.0	52.0	9.0	24.0	B	4.0	65.0	6.0	15.0
17	B	5.0	46.0	16.0	23.0	B	19.0	54.0	1.0	16.0
18	C	9.0	4.0	65.0	12.0	D	6.0	2.0	27.0	55.0
19	B	10.0	77.0	1.0	2.0	C	0.0	7.0	79.0	4.0
20	C	29.0	29.0	25.0	7.0	D	52.0	7.0	7.0	24.0
21	B	2.0	26.0	62.0	0.0	A	28.0	6.0	55.0	1.0
22	A	38.0	31.0	16.0	5.0	B	2.0	50.0	30.0	8.0
23	D	28.0	17.0	4.0	41.0	B	21.0	42.0	5.0	21.0
24	B	20.0	63.0	2.0	5.0	B	14.0	68.0	1.0	7.0
25	C	8.0	18.0	46.0	18.0	C	8.0	16.0	57.0	9.0
26	B	25.0	48.0	14.0	3.0	B	27.0	54.0	7.0	2.0
27	C	16.0	11.0	50.0	13.0	C	11.0	1.0	77.0	1.0
28	A	31.0	46.0	7.0	6.0	A	53.0	29.0	5.0	3.0
29	C	23.0	5.0	55.0	7.0	B	47.0	38.0	3.0	2.0
30	A	75.0	4.0	11.0	0.0	A	87.0	0.0	2.0	1.0