UNIVERSITY OF SOUTHAMPTON

FACULTY OF LAW, ARTS & SOCIAL SCIENCES

School of Education

EFL Students' English Language Knowledge, Strategy Use and Multiple-choice Reading Test Performance: A Structural Equation Modeling Approach

by

Wei-Tsung Hsu

Thesis for the Degree of Doctor of Philosophy

June 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF LAW, ARTS & SOCIAL SCIENCES
SCHOOL OF EDUCATION

Doctor of Philosophy

EFL STUDENTS' ENGLISH LANGUAGE KNOWLEDGE, STRATEGY USE AND MULTIPLE-CHOICE READING TEST PERFORMANCE: A STRUCTURAL EQUATION MODELING APPROACH

In Taiwan, a reading comprehension component is included in the English test of the Senior High Academic Ability Examination (SHAAE) – a national examination which can be regarded as a university entrance examination for students in their final year of senior high. This reading subtest consists of a multiple-choice format. Studies on language assessment, L2 reading and L1-L2 reading have suggested that EFL students' performance on multiple-choice reading comprehension tests is attributed to two major factors: English Language Knowledge and Strategy Use. This feature raises a number of issues. Does the multiple-choice reading comprehension subtest of the English component at the SHAAE measure what it is intended to assess? Do Taiwanese senior high school students' English Language Knowledge and Strategy Use have an effect on their multiple-choice reading comprehension test performance? What are the relative contributions of students' English Language Knowledge and Strategy Use to their reading comprehension test performance? Is there a language threshold for students' deploying some strategies to contribute to their reading test performance? The current study sets out to address these issues. It investigates the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance. The findings of the research are connected with: (a) the English language teaching approach for English language teachers in Taiwan; (b) the validity of the reading comprehension subtest of the English component at the SHAAE; and (c) the validity of salient models of language ability.

A quantitative research approach is used that involves an ex post-facto correlational research design, utilizing survey methodology. An English Language Knowledge test, a Strategy Use questionnaire, and a multiple-choice reading comprehension test serve as instruments. 1064 EFL students in six senior high schools located in the south region of Taiwan participated in the study. Data was collected in the classroom during English class sessions. Participants took a reading test and completed a Strategy Use questionnaire. Three to seven days later, they sat an English Language Knowledge test. Exploratory factor analysis is conducted to extract components underlying the data collected from instruments. Structural Equation Modeling is applied to examine the relationship among students' English Language Knowledge, Strategy Use and their reading test performance.

The main finding of the study is that Taiwanese senior high school students are strategic readers/test-takers. Their English Language Knowledge and Strategy Use contribute to their reading test performance. However, compared with that of English Language Knowledge, the contribution of students' Strategy Use to their reading test performance is smaller. In addition, a language threshold is present for students deploying strategies contributing to their reading test performance. In conclusion, the thesis addresses the need for implementing strategy instruction for students to improve their Strategy Use in a reading test and further to promote their reading test performance. The discussion also compares the outcome of the research with other approaches to Reading/Test-taking Strategy Use and current models of Strategic Competence.

CONTENTS

CHAPTER FIVE: MULTIPLE GROUP ANALYSES: RESULTS

CHAPTER SIX: DISCUSSION

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I,…………………...................**Wei-Tsung Hsu**…………………………………………......,

declare that the thesis entitled

EFL Students' English Language Knowledge, Strategy Use and Multiple-choice Reading Test Performance: A Structural Equation Modeling Approach……………..

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed: *Wei-Tsung Hsu*

Date: 11 June, 2008

ACKNOWLEDGEMENTS

As my research study is coming to an end, I often feel that I am fortunate to have obtained the help from many people who not merely assist in setting me on the appropriate path but aid me to overcome apparently insurmountable obstacles in the course of the study. There are several people I wish to thank.

First, I would like to thank my supervisors, Professor Mike Grenfell and Dr. John Schulz, for their patient guidance and invaluable suggestions throughout my research. But for them, this thesis would not be completed.

Additionally, I wish to express my gratitude to Dr. Sheng-hong Chen for his response to my questions concerning statistics and help to get reference books about AMOS.

Moreover, I am greatly grateful to Dr. Ru-gang Lu, Dr. Wen-hsien Yang and Ms. Mei-chueh Chao for their inspecting my measuring instruments.

In addition, I would like to thank Ms. Yi-chun Feng and Ms. Te-yuan Hsu for their examinations of the Chinese version of my questionnaire.

I am also indebted to students and English teachers in six senior high schools in the south region of Taiwan. It was through them that the relevant data could be gathered and then my research questions could be answered. Given the number of students and teachers and the protection of their anonymity, I do not mention all their names.

Finally, I sincerely appreciate my family's full support and contribution to the fulfillment of this study.

List of Abbreviations

CFA	Confirmatory factor analysis
CMDRS**	Constructing the meaning with the deployment of reading strategies – a strategy subgroup [an observed variable for constructing the meaning and evaluating (CME)]
CME*	Constructing the meaning and evaluating – a strategy use process (a latent variable)
EFA	Exploratory factor analysis
EFL	English as a foreign language
ELK	English language knowledge
ELKT	An English language knowledge test
EM*	Evaluating and marking – a strategy use process (a latent variable)
EVA**	Evaluating – a strategy subgroup [an observed variable for constructing the meaning and evaluating (CME)]
ExQ**	Explicit questions – a reading test item subgroup [an observed variable for multiple-choice reading comprehension test performance (MC RCTP)]
FL	Foreign language
GK*	Grammatical knowledge (a latent variable)
GRAM1**	GRAM1 consists of eight test items of the grammar subtest – a grammatical-knowledge test item subgroup [an observed variable for grammatical knowledge (GK)]
GRAM2**	GRAM2 consists of eight test items of the grammar subtest – a grammatical-knowledge test item subgroup [an observed variable for grammatical knowledge (GK)]
HEA	High English Ability
II**	Interacting with the input – a strategy subgroup [an observed variable for constructing the meaning and evaluating (CME)]
InQ**	Inferential questions – a reading test item subgroup [an observed variable for multiple-choice reading comprehension test performance (MC RCTP)]
L1	First language
L2	Second language
LEA	Low English ability
LEX1**	LEX1 consists of ten test items of the vocabulary subtest – a lexical-knowledge test item subgroup [an observed variable for lexical knowledge (LK)]
LEX2**	LEX2 consists of eleven test items of the vocabulary subtest – a lexical-knowledge test item subgroup [an observed variable for lexical knowledge (LK)]
LK*	Lexical knowledge (a latent variable)
MC RCT	A multiple-choice reading comprehension test
MC RCTP*	Multiple-choice reading comprehension test performance (a latent variable)
MDAMT*	Monitoring, directing attention and managing the test – a strategy use process (a latent variable)
MIP**	Marking incomprehensible parts – a strategy subgroup [an observed variable for evaluating and marking (EM)]
MKPO**	Marking key points or options – a strategy subgroup [an observed variable for evaluating and marking (EM)]

MRPNR**	Monitoring the reading process with negative results – a strategy subgroup [an observed variable for monitoring, directing attention and managing the test (MDAMT)]
MRPPR**	Monitoring the reading process with positive results – a strategy subgroup [an observed variable for monitoring, directing attention and managing the test (MDAMT)]
MTDTS**	Managing the test with the deployment of test-taking strategies – a strategy subgroup [an observed variable for monitoring, directing attention and managing the test (MDAMT)]
MTTP**	Monitoring the test-taking process – a strategy subgroup [an observed variable for monitoring and utilizing test questions (MUTQ)]
MUTQ*	Monitoring and utilizing test questions – a strategy use process (a latent variable)
REP**	Repeating – a strategy subgroup [an observed variable for monitoring, directing attention and managing the test (MDAMT)]
RL**	Retrieving-linking – a strategy subgroup [an observed variable for monitoring, directing attention and managing the test (MDAMT)]
RTSU	Reading and test-taking strategy use
SEM	Structural equation modeling
SHAAE	Senior High Academic Ability Examination
SHASE	Senior High Appointed Subject Examination
TATQ**	Taking advantage of test questions – a strategy subgroup [an observed variable for monitoring and utilizing test questions (MUTQ)]

Note. * represents a latent variable, while ** an observed variable in the SEM analysis in the current study.

CHAPTER ONE
INTRODUCTION

1.1 Introduction

This study has two key objectives. The first is to investigate the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and performance on reading comprehension tests. The second is to examine whether this relationship differs across English ability levels.

To conduct this study, I limited *knowledge of the English language* to lexical and grammatical knowledge. *Lexical* knowledge denotes students' breadth of vocabulary, whereas *grammatical* knowledge signifies students' knowledge of syntactic rules, prepositions and word usage. *Reading and test-taking strategies* refer to the conscious and/or subconscious mental and behavioral activities that affect student performance on multiple-choice reading comprehension tests – either directly or indirectly. *Reading and test-taking strategy use* relates to the deployment of these strategies. *Multiple-choice reading comprehension test performance* refers to how well students perform on a multiple-choice reading comprehension test that measures their ability to read for main ideas, facts, and details of particular reading passages. It also refers to how well they draw inferences.

I begin Chapter I by describing the educational system in Taiwan. I then provide the background and purpose of this study and pose several basic research questions. Finally, I explain the significance of the study and present a general outline for the rest of this thesis.

1.2 Educational system in Taiwan

To put this study into context, it is important to understand the basic structure of the educational system in Taiwan – from elementary school through high school. Taiwanese students start elementary school at the age of seven or eight. After six years, they move to junior high school for three years. Both elementary school and junior high school are compulsory. If junior high school graduates want to continue their education, they take the Junior High Basic Academic Ability Examination. Depending on the scores they receive in this test and on their interests, they then attend either senior high school or a vocational school for three more years. Senior high school graduates who want to attend university must sit for – and pass – either the Senior High Academic Ability Examination (SHAAE) or the Senior High Appointed Subject Examination (SHASE). Vocational school graduates who want to study further must take either a two-year or a four-year College Entrance Examination.

1.3 Background and motivation for the study

With the rise of globalization and the revolution in information technology, the role that English plays in Taiwan's daily life is increasingly important. To improve English language proficiency and accelerate the ability to communicate with members of the global village, students begin English language instruction in the third grade. The Ministry of Education has developed a curriculum for English instruction at the elementary level that focuses mainly on the development of students' English listening and speaking skills, although reading and writing skills are also developed (Department of Elementary Education in Taiwan, 2008). The curriculum for junior and senior high school students centers on the equal development of all four basic skills: listening, speaking, reading and writing (Department of Elementary Education in Taiwan, 2008; English Education Resource Center in Taiwan, 2008). Students at this level study English as a school subject and often take three to five English classes per week.

Although the curriculum for English instruction in senior high school stresses equal development of listening, speaking, reading and writing skills, most teachers actually give more attention to reading than they do to any other skill. In particular, they spend a large amount of class time helping students make sense of the material in textbooks. One reason for this may be the fact that senior high school graduates need to acquire a high level of proficiency in reading in order to understand textbooks and academic journals in university. The most important reason, however, is that reading is the most predominant skill measured on the English portion of a critical examination – the Senior High Academic Ability Examination (SHAAE). The subtests that assess reading ability comprise 70% of the English test. (To see a sample of the English test, go to http://www.ceec.edu.tw/AbilityExam/ AbilityExam Paper.htm). This means that the ability to read English plays a major role in students' ability to achieve high scores on the English test of the SHAAE.

The Senior High Academic Ability Examination (SHAAE), held in February, is a national exam for third-year senior high school students. This exam, which consists mainly of multiple-choice test items, assesses what students learned in the first and second years of high school in five subjects: Chinese, English, mathematics, science and social science. With the examination scores, students can apply to a university. Admissions committees in each university use the SHAAE as a preliminary criterion with which to choose (or eliminate) students for the second stage of the selection process. Students who are rejected in this first round can take another, more challenging examination in July called the Senior High Appointed Subject Examination (SHASE). Doing well on this exam makes it possible for them to enter university. In order to avoid taking this examination and to have a greater chance of attending university, almost all senior high third-graders take the SHAAE. As a result, the outcome of the SHAAE is vital for high school students.

The reading comprehension subtest is included in an English test of the SHAAE. This subtest is in a multiple-choice format. Language assessment studies have demonstrated that test-takers' cognitive processes differ to some extent between regular reading contexts and contexts in which multiple-choice reading comprehension tests are taken (e.g., Gordon & Hanauer, 1995; Rupp, Ferne, & Choi, 2006). Test-takers clearly understand that multiple-choice reading tests require a different approach from "normal" circumstances; consequently, they are eager to take advantage of test-taking strategies to better their performance on such tests (Cohen, 1984; Cohen & Upton, 2006; 2007; Nevo, 1989; Rupp, Ferne, & Choi, 2006). In addition, studies have indicated that strategy deployment varies with test items designed to measure disparate facets of reading comprehension (e.g., Anderson, Bachman, Perkins, & Cohen, 1991). Studies have also shown that test-takers frequently use matching strategies in multiple-choice reading tests and more frequently in L2[1] reading tests (e.g., Farr, Pritchard, & Smitten, 1990; Nevo, 1989). Given the critical importance of the SHAAE results to Taiwanese high school students, it is important to ask the question: Does the reading comprehension subtest of the English component actually measure what it is intended to measure?

In addition to what is stated above, another reason I undertook this study grew out of my experience in teaching English in Taiwan and of three questions I had. My students used to complain frequently about the difficulty of the English reading comprehension subtest of the SHAAE. My colleagues who were teaching in other schools had similar experiences. Literature on L2 reading has indicated that both lexical knowledge and grammatical knowledge are related to, and even exert an effect on, L2 reading test

[1] This study adopts a broad definition of an L2 (a second language); that is, it includes both EFL (English as a foreign language) and ESL (English as a second language).

performance (e.g., Kobayashi, 2002; Purpura, 1997; 1998b; 1999; Shiotsu & Weir, 2007; Taillefer, 1996; Usó-Juan, 2006). Clearly, improving students' overall knowledge of the English language solves this problem. In addition to this, can anything else be done to help them?

Studies on reading strategies have found that L2 readers use a variety of reading strategies to overcome obstacles to their comprehension when processing L2 texts (e.g., Block, 1986; 1992; Hosenfeld, 1984; Yang, 2002; 2006). Readers who are metacognitively aware of their reading process invoke strategies appropriately and flexibly, which then further promotes their reading task performance (Jiménez, García, & Pearson, 1996; Yang, 2006). Furthermore, studies have also found that the strategies of L2 readers with high proficiency vary to a certain degree from those of L2 readers with low proficiency (e.g., Cziko, 1980; McLeod & McLaughlin, 1986; Upton & Lee-Thompson, 2001). Successful readers are more meaning-centered and top-down oriented in their strategy employment (Block, 1992; Devine, 1984; Sheorey & Mokhtari, 2001; Yiğter, Sariçoban, & Gürses, 2005). In contrast, less successful readers tend to deploy bottom-up, or negative, strategies to solve comprehension breakdowns in their reading (Block, 1992; Hosenfeld, 1984; Knight, Padron, & Waxman, 1985; Yamashita, 2002). They are also more sound-centered and word-based and possibly use more local strategies (Auerbach & Paxton, 1997; Devine, 1984). Finally, studies have illustrated that students can improve their reading performance through strategy instruction (e.g., Auerbach & Paxton, 1997; Barnett, 1988; Carrell, Pharis, & Liberto, 1989; Farrell, 2001; Kern, 1989; Macaro & Erler, 2008).

It appears from the studies mentioned above that Taiwanese high school students would benefit greatly from being trained in how to deploy reading and test-taking strategies appropriately on the reading comprehension subtest of the English component at the SHAAE. Prior to commencing such instruction, however, it is important to answer three questions. The first is: Do the reading strategies that Taiwanese senior high school students use affect their performance on multiple-choice reading comprehension tests in English? If the answer is "Yes," then it is important to discover the size of this effect. This would enable us to understand how students currently approach multiple-choice reading comprehension tests; it would also serve as a frame of reference for implementing strategy instruction in the future. A qualitative study could identify the effects of strategy use on reading test performance. This is based on evidence that readers do invoke strategies to deal with the parts of an L2 text they do not understand and that these parts can be solved through the use of strategies. Nevertheless, little information is shown about the strength of the effect that readers' strategy use yields on their reading test performance. Hence, to answer this question, I have adopted a quantitative-dominated research approach.

The second question that must be answered before commencing instruction in reading and test-taking strategies is the relative contributions of students' English language knowledge and strategy use to their multiple-choice reading comprehension test performance. The answer to this question is related to which components should be prioritized in English classes: English language knowledge or strategy use? L1-L2 reading research indicates that L2 proficiency or language knowledge has a greater influence on L2 reading performance than L1 reading ability (e.g., Bernhardt & Kamil, 1995; Bossers, 1991; Lee & Schallert, 1997; Taillefer, 1996; Yamashita, 2002). Most previous studies in this area have investigated L1 reading ability rather than strategy use, so little research is available that helps to answer this question.

The third question is: Does a language threshold exist for students' ability to use strategies on multiple-choice English reading comprehension tests? In other words, do students need to reach a certain level of the knowledge of English in order to successfully apply reading and test-taking strategies to multiple choice tests? The answer to the question can provide more insights into the role that English language knowledge plays in students' strategy deployment, which is also associated with strategy instruction. Previous studies have suggested the presence of a language threshold for transferring L1 reading ability to L2 reading (e.g., Bernhardt & Kamil, 1995; Bossers, 1991; Clarke, 1980; Lee & Schallert, 1997; Taillefer, 1996; Yamashita, 2002). Clearly, L2 readers do need to gain a certain level of L2 proficiency before they are able to apply their L1 reading ability to their L2 reading. However, in most L1-L2 reading studies, L1 reading ability is measured by an L1 reading test. The data derived from such studies is unrelated to strategy use. Therefore, a limited amount of empirical data is available to answer my third question.

The aforementioned three questions are partially responsible for my conducting the current study. Bachman and Palmer (1996) put forward a model (see Section 2.8.2) of language ability in language use and language test performance that operationalized strategic competence using an array of metacognitive strategies (e.g., planning and goal setting). However, only a limited amount of research has been carried out to validate their model and to address the construct of strategic competence (e.g., Nikolov, 2006; Phakiti, 2003; 2008; Purpura, 1997; 1999). As several language assessment researchers suggest (e.g., Nikolov, 2006; Purpura, 1999), more research is still needed to provide more empirical evidence for Bachman and Palmer's (1996) model of language ability in language use and language test performance.

Given the questions stated above and some language testing researchers' suggestion, I undertook the current study to investigate the relationship among Taiwanese senior high school students' English language knowledge, strategy use, and performance on multiple-choice reading comprehension tests.

Related to the current study are Purpura's (1997; 1998b; 1999) research works. His studies, with the use of structural equation modeling (SEM), provide empirical evidence for (a) strategy use having an effect on L2 test performance; (b) a language threshold being present for strategy deployment to have an effect on L2 test performance; (c) strategy use differing between the high English ability (HEA) group and the low English ability (LEA) group to some extent.

However, several drawbacks are present in his research. To begin with, because the participants in his studies did not refer to given tasks when they filled in the strategy use questionnaire, the collected data may be somewhat unreliable. Further, Purpura adopted participants' L2 test results, which were involved in parameter estimation in the SEM analysis, to divide participants into the HEA group and the LEA group. Such manipulation makes it easy to manifest cross-group differences in test performance in individual group models. For example, a grammar subtest assessed lexico-grammatical ability much better in the HEA group (with a factor loading of .577) than in the LEA group (with a factor loading of .186). Therefore, some identified group differences found in the study are questionable. In addition, Purpura focused on the relationship between strategy use and performance on an L2 test (reading, vocabulary and grammar tests); thus, the results provided limited insights into the relationship among L2 language knowledge, strategy use, and reading test performance. Purpura's L1 participants were also heterogeneous, as was their course level. These variances may impact participants' strategy use and L2 test performance.

Taking into consideration the shortcomings of Purpura's studies, I am interested in discovering whether similar findings will be produced in a study where (a) participants' L1 and course levels are homogeneous; (b) participants' strategy use is elicited through the presence of a task; (c) a different criterion is adopted for group division; and (d) the research focus is on English language knowledge, strategy use, and reading test performance.

1.4 Purpose of the study

The purpose of the present study is twofold. The first is to investigate the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and performance on multiple-choice reading comprehension tests. The goal is to obtain definitive answers to the questions: Do the degree of knowledge of the English language and students' ability to successfully apply reading and test-taking strategies influence their performance on multiple-choice reading comprehension tests and how? The answers to these questions will contribute to teachers' understanding of how they can better prepare their students to do well on the challenging reading comprehension subtest included in an English test of the critically-important SHAAE. It will also assist them (and the examination center in Taiwan) to better understand what the reading comprehension test scores actually mean. In addition, armed with such information, English language teachers in Taiwan will be able to decide whether or not to teach reading and test-taking strategies and when they should teach them.

The second purpose of this study is to examine whether the relationship among English language knowledge, reading and test-taking strategy use, and performance on multiple-choice reading comprehension tests differs at varying levels of ability in English. The goal is to pinpoint cross-group commonalities and differences in the way that students' knowledge of English and their use of reading and test-taking strategies affect their performance on multiple-choice reading comprehension tests, and the size of such effects. The answer to this question will help teachers in Taiwan implement strategy instruction that improves students' scores on multiple-choice reading tests of English.

1.5 Research questions

The current study will attempt to answer the following research questions:

1. What is the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance?
   1.1 Do students' English language knowledge and reading and test-taking strategy use contribute to their multiple-choice reading comprehension test performance? If yes, what are their relative contributions to multiple-choice

reading comprehension test performance?

1.2 Is there a language threshold for students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance?

1.3 What is the relationship between students' English language knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test?

2. Is there a difference in the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance across English ability levels?

   2.1 Is there a difference in students' English language knowledge and reading and test-taking strategy use contributing to their multiple-choice reading comprehension test performance across English ability levels? Do the relative contributions of students' English language knowledge and reading and test-taking strategy use to their multiple-choice reading comprehension test performance differ across English ability levels?

   2.2 Is there a language threshold for students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance across English ability levels?

   2.3 Is there a difference in the relationship between students' English language knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test across English ability levels?

In Chapter Four, I discuss the process used to analyze the first research question and it sub-questions, and present the results. In Chapter Five, I discuss the process and results for the second research question. I present the answers to these research questions in Chapter Six.

1.6 Significance of the study

This exploration of the relationship among Taiwanese senior high school students' English language knowledge, use of reading and test-taking strategies, and performance on multiple-choice reading comprehension tests produces a number of findings and implications that contribute to the pedagogy of English language instruction in Taiwan. It also makes theoretical and methodological contributions to the research fields of L2 reading, strategy use and language testing.

Pedagogically, this study supplies Taiwanese high school students and English language teachers with valuable information regarding the ways in which English language knowledge and use of reading and test-taking strategies affect performance on multiple-choice reading tests. Such knowledge gives teachers a better understanding of which components they should put more emphasis on in English classes, English language knowledge or strategy use, at different stages of learning. This study also provides helpful insights into the reading and test-taking strategies that Taiwanese high school students use (or don't use) on multiple-choice reading tests. These findings will enable Taiwanese English language teachers to create a frame of reference with which to improve their students' knowledge of reading and test-taking strategies and their ability to apply this knowledge on multiple choice reading tests.

From a theoretical perspective, this study provides more empirical evidence for Bachman's (1990) factors which impact upon test scores, and for Bachman and Palmer's (1996) model of language ability. Unlike most previous studies (e.g., Nikolov, 2006; Phakiti, 2003) in this area, the evidence is predicated on the presence of the effects of English language knowledge and strategy deployment on reading test performance in a single modeling framework that profiles the relation among English language knowledge, strategy use, and reading test performance. This study also gives empirical evidence regarding the influence of language thresholds on strategy use and thereby on reading test performance. Distinct from most previous studies (Bernhardt & Kamil, 1995; Bossers, 1991; Lee & Schallert, 1997), the evidence is based on data collected by means of a questionnaire on strategy use, not on L1 reading ability data gathered by an L1 reading test.

This study also provides useful insights into how EFL students' use of reading and test-taking strategies impacts their performance on reading comprehension tests and how their deployment of these strategies varies according to the level of English ability. In contrast to most other studies (e.g., Cziko, 1980; McLeod & McLaughlin, 1986; Upton & Lee-Thompson, 2001), however, this study investigates the variations by modeling and testing the relationship among English language knowledge, strategy use, and reading test performance across groups with different levels of English ability.

Methodologically, the structural equation modeling (SEM) approach is applied in the current study to examine the relationship among English language knowledge, strategy use, and reading test performance. SEM is a multivariate analytic procedure for examining inter-relationships among a set of variables of interest. It allows an effect of a variable on another to be shown in a single modeling framework. Until now, only a handful of studies (e.g., Phakiti, 2008; Purpura, 1997; 1998b; 1999) have conducted SEM to investigate the relation between strategy deployment and L2 test performance.

Further, this study uses SEM to conduct multiple-group analyses that create both a high English ability group model and a low English ability group model. This makes it possible to locate cross-group commonalities and discrepancies between the two models. It also enables to estimate the two group models simultaneously with equality constraints imposed on parameters of interest to provide more robust evidence for cross-group commonalities and variations. Until now, simultaneous group analysis has been performed on a limited number of studies (e.g., Purpura, 1998b; 1999) that explore the relationship between strategy employment and L2 test performance. From an alternative perspective of data analysis, the findings yielded in this study can confirm and/or disconfirm those found in previous qualitative and quantitative studies.

1.7 Structure of the thesis

This thesis is divided into seven chapters. Chapter One describes the educational system in Taiwan, gives the background and motivation for this study, presents the research questions, and discusses the overall purpose and significance of the study.

Chapter Two discusses the theoretical framework used in the study and reviews the relevant literature. This includes reading strategies/processes, test-taking strategies/processes, a language threshold for transferring L1 reading ability/strategies to L2 reading, and so forth. It also discusses the limitations of previous studies.

Chapter Three discusses the methodology used in this paper, including the surveys and measurements. Topics covered include research design; study participants; and data collection techniques, procedures, and analysis methods. It also describes the pilot study.

Chapter Four describes the results of modeling the relationship among English language knowledge, use of reading and test-taking strategies, and performance on multiple-choice reading comprehension tests. It relates the results to the first research question and its sub-questions. In addition, it provides a brief discussion of these results.

Chapter Five describes the results of the multiple-group analyses. More specifically, it presents the results of the separate group analysis and the simultaneous group analysis. It also relates this information to the second research question and its sub-questions and discusses it briefly.

Chapter Six discusses the major findings. It gives the answers to both of the research questions and their sub-questions and compares the findings of this study with those of other studies.

Chapter Seven presents the implications of this study for the college entrance examination center in Taiwan and for English language teachers at the senior high school level in Taiwan. This chapter also addresses the limitations of this study and gives recommendations for further research.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

In Chapter Two, I present a literature review that gives readers some background knowledge of the research questions investigated in this study. The theoretical framework is twofold: L2 and L1-L2 research on the fields of reading and language testing. The chapter begins with an overview of reading models and language learner strategies. It then moves to studies that focus on L2 reading strategies, followed by a review of studies on (a) the roles that lexical knowledge and grammatical knowledge play in L2 reading; (b) L2 language knowledge and strategy use in L2 reading; and (c) a language threshold for transferring L1 reading ability/strategies to L2 reading. In particular, I discuss factors proposed by Bachman (1990) that influence test scores and Bachman and Palmer's (1996) model of language ability. Finally, I review studies related to the use of strategies for multiple-choice reading tests.

2.2 Reading models

A great deal of research has already been conducted on the L1 reading process and models of the reading process have been constructed. By drawing on the linguistic interdependence hypothesis, scholars have attempted to apply the insights from L1 reading research to L2 (or FL) reading (Erler & Finkbeiner, 2007; Grabe & Stoller, 2002). Some limitations are present, however, since L2 reading encompasses such L1 influences as L1 reading ability and L1 language knowledge, and is consequently much more complex. Despite the drawbacks, reading models gained from L1 reading do contribute to our better understanding of how readers read L2 texts. In the following, models of the reading process – *bottom-up*, *top-down* and *interactive* models – are briefly discussed.

Until the late 1960s, research on reading and reading instruction mainly followed a *bottom-up* model (Parry, 1996). Also called the data-driven or text-driven approach, this model conceived of reading as a linear process that entailed decoding written symbols into their aural equivalents. Readers, at first, discriminate each letter that they encounter in texts, decode these letters to sound, match the written symbols with their aural equivalents, integrate these to form words, and finally derive the meanings of the words. With the repetition of the aforementioned overall procedures, and the assistance of their long-term memory of background knowledge, readers gradually construct an interpretation of the entire text. In this model of reading, each component operates independently of one another and builds upon the antecedent component (Alderson, 2000). Theories of reading viewing “the use of an intermediate speech code (i.e., any form of phonemic recording) as an essential process in reading competence…are usually considered bottom-up views of reading” (Cziko, 1980: 101).

In the late 1960s and 1970s, researchers began to propose an alternative model called *top-down* processing (e.g., Goodman, 1967; Smith, 1982). Also called the concept-driven or reader-driven approach, this model recognizes the critical role played by readers’ expectations of the contents of the text being processed (Urquhart & Weir, 1998). This perspective regards reading as a *psycholinguistic guessing game* in which readers equipped with different schemata, or structures of knowledge, commence with a host of predictions, hypotheses, or expectations about the meaning expressed by the text they are going to read. They then sample the text to confirm or reject the previous predictions. In this approach, reading is a process of reconstructing the overall meaning of texts. It downplays the significance that reading texts themselves conventionally bear and emphasizes the importance of what readers themselves bring to the process (Alderson, 2000). Samuels and Kamil (1988) point out that a discrepancy between bottom-up and top-down models lies in that “the bottom-up models start with the printed stimuli and work their way up to the higher-level stages, whereas the top-down models start with hypotheses and predictions and attempt to verify them by working down to the printed stimuli” (p. 31).

Bottom-up and top-down approaches competed with each other throughout the 1970s and into the 1980s until a general consensus began to emerge that reading is a complicated, interactive process that involves both approaches (Carrell, 1988; Stanvich, 1980; Urquhart & Weir, 1998). As Wolff (1987) puts it, both data-driven processing and concept-driven processing are “interdependent processes” (p. 311). Each compliments the other. It is difficult to process L2 reading texts on the strength of either the top-down or the bottom-up approach alone. The interactive approach of reading, on the whole, consists of dual notions (Grabe, 1991; Grabe & Stoller, 2002). The first denotes that an interaction occurs between readers and texts. To construct meaning, readers must draw on both what they see in the text and on their prior knowledge (Bernhardt, 1991; 2005; Carrell, 1988; Carrell & Eisterhold, 1983; Nassaji, 2002). The second signifies that multiple components interact with each other simultaneously in the reading process – from a low-level skill such as word recognition to a high-level skill such as synthesis or evaluation (Koda, 2005; Nassaji, 2003b; Samuels & Kamil, 1988; Stanvich, 1980). Grabe (1991) notes that "these two perspectives are complementary" (p. 383).

In a regular setting, readers constantly alternate between bottom-up and top-down approaches. They may start with bottom-up reading to process a chunk of a sentence and then shift to top-down reading to make a hypothesis about the meaning of a sentence or a group of sentences. They will use top-down reading to predict the meaning of the input, then switching to bottom-up reading to check whether their prediction is correct. The interactive approach, obviously, serves as a felicitous way to profile the actual information processing pertinent to reading. In such reading processing, readers have to process texts in a strategic manner given a need to perform myriad sub-processes such as word recognition or syntax parsing simultaneously and the limited processing capacity, or the potential presence of hindrances to conducting the sub-processes well. It follows that the importance of strategy deployment and metacognitive awareness cannot be overemphasized.

2.3 Language learner strategies

In the field of language learning, the last four decades have witnessed numerous applied linguists, cognitive psychologists, and educational psychologists devote themselves to research on learning strategies in the attempt to uncover the mental processes that benefit individual learning. We can characterize a strategy as purposeful, essential, effortful, procedural, willful and facilitative. As such it can be viewed as representative of procedural knowledge that refers to the "how to" knowledge with which learners are equipped (Alexander, Graham, & Harris, 1998). Such knowledge functions as a frame of reference that learners count on to surmount obstacles to their learning or to boost their performance on given tasks.

According to Cohen (1998b), within the L2 context, language learner strategies can be classified into two categories: *language learning* and *language use*. Language learning strategies are those that language learners draw upon to promote language learning and acquisition in general (Phakiti, 2003). In other words, language learners use these strategies to facilitate the acquisition of overall knowledge and skills, usually in a normal situation. For example, language learners read English newspapers every day to enhance their English reading ability. By contrast, language use strategies are those that language learners use to successfully achieve their goals in a specific context (e.g., to obtain better scores on a reading test in a time-constrained test setting). As Phakiti (2003) puts it, language learning strategies can be regarded as continuing and incessant activities, in contrast to language use strategies which are setting-oriented.

While related to a specific language skill domain, language learner strategies are named according to that skill (Oxford *et al*., 2004). Consequently, different derivatives emerge such as reading strategies, or listening strategies. For L2 reading strategy researchers, in a particular context in which a task is provided, language use strategies probably have been what they are interested in, since generally it is these strategies, rather than language learning strategies, are elicited when L2 readers are on task. On the other hand, as connected with test settings, language learner strategies can be labeled as test-taking strategies. Similar to some L2 reading strategy researchers, language assessment researchers might be more concerned about language use strategies in that it is these strategies, not language learning strategies, that have a direct impact on test-takers' test performance.

Cohen's (1998b) classification of language learner strategies functions as an indication for a significant development in language learner strategy research (Anderson, 2005). Using this classification as a reference point, researchers have an understanding of what types of strategies their studies focus on. However, as yet "no research has been conducted…to determine if this categorization of strategies is valid" (Anderson, 2005: 762). Hsiao and Oxford (2002) comment that "both learning and use can occur simultaneously; and in daily reality the strategies for L2 learning and L2 use overlap considerably" (pp. 378-379). It appears challenging to draw a precise distinction between language learning strategies and language use strategies. The difficulty in categorizing strategies in a clear-cut manner is further illustrated as follows.

Rubin (1981), and O'Malley and Chamot (1990) identified language learner strategies, such as deductive reasoning and transferring, that can be used in different content areas, such as English or math. Alexander et al. (1988) referred to these as "general cognitive strategies" (p. 132). This type of strategy is distinct from task-specific strategies, which are restricted to a certain task (e.g., a multiple-choice reading comprehension test). That is, they are bounded (p. 32). The reading and test-taking strategies used in multiple-choice reading comprehension tests can be affiliated with this form of strategy, on the grounds that they are drawn upon in the context where multiple-choice reading comprehension tests are taken. However, the real case is not that simple. Some strategies (e.g., test-takers try to consult options to obtain some related information) deployed in multiple-choice reading comprehension tests can be employed across different tasks (e.g., multiple-choice listening comprehension tests). In other words, they are not task-specific. This indicates the difficulty in classifying strategies categorically. Such a difficulty may be pertinent to the fuzziness of the definition of strategies.

In the realm of language learner strategies, the definition of strategies has been a debatable issue. As for the degree of consciousness involved in strategies, some strategy researchers claim that strategies are referred to as activities or behaviors deployed consciously, subconsciously or unconsciously (e.g., Barnett, 1988; Faerch & Kasper, 1983; Kern, 1989). On the other hand, some argue that only activities or behaviors employed in a conscious way can be looked on as strategies (e.g., Anderson, 2005; Cohen, 1998b, Ellis, 1994; Pritchard, 1990; Williams & Moran, 1989). Despite the debate of the clear-cut extent of consciousness, most researchers (i.e., Chamot, 2005; Cohen, 1998b; Hsiao & Oxford, 2002; Oxford & Cohen, 1992) agree that the involvement of a certain level of conscious intention is an indispensable element in employing strategies.

Another controversial issue is concerned with whether strategies are mental operations or behavioral activities. Some strategy researchers view strategies as mental operations that language learners deploy in L2 acquisition, L2 use or L2 test contexts (e.g., Abbott, 2006; Cohen, 1998b; Hosenfeld, 1977; Macaro, 2006). At the same time, among several researchers (e.g., Anderson, 2005; Ellis, 1994; Purpura, 1997; 1998b; 1999), "there remained [remains] a determination that strategies should encompass more than mental operations" (Grenfell & Macaro, 2007: 21). That is, strategies are conceived as both mental and behavioral activities related to given task performance. It is clear that for the strategy definition, a precise distinction is not supplied between overt motor behavior and mental activity (Macaro & Erler, 2008).

Macaro (2006) notes that the definition of strategies "in the literature is arrived through the use of equally undefined terms" (p. 325). The controversial definition of strategies seems associated with methods utilized to investigate strategies that learners deploy. As verbal reports are applied to examine learners' strategy deployment, it appears evident that actions or activities occurring consciously will be verbalized and detected. On the other hand, when questionnaires are adopted, subconscious or potentially unconscious activities, in addition to conscious ones, may be self-reported. In this study, a strategy is defined as a consciously or subconsciously, mental or behavioral activity related directly or indirectly to task performance.

The techniques utilized to collect the strategy data is also related to the approaches adopted in the research. When the focus is on understanding the strategy use of a large group of language learners, self-reported questionnaires will be employed, particularly structured questionnaires to collect data. The data allows inferential analysis to be conducted and a quantitative research approach is adopted. Alternatively, when the concern is on deeply examining learners' strategy use on a given task or understanding the development of learners' strategies from a small sample, interviews, verbal reports, diaries and journals tend to be utilized to gather data. Usually the "thick" and "raw" data is obtained and transcribing the data is necessary. Then, it is a qualitative research approach that is adopted, although sometimes the data is quantified and some statistics are performed (e.g., count the frequency or calculate the mean). In the following section, a focus will shift to reading strategies/processes related to L2 reading, since the current study examined Taiwanese senior high school students' strategy use in a reading test.

2.4 Reading strategies/processes related to L2 reading

Within the field of reading, great attention has been drawn to investigating how readers interact with texts in the reading process. A number of factors that influence the nature of reading such as text organization, readers' strategy use, readers' language knowledge have been pinpointed – Alderson (2000) provides details for these factors. With regard to reader-based factors, reading research uncovers readers' black box and casts light on the concept that readers' characteristics affect reading performance. Being equipped with divergent purposes, interests, attitudes or background knowledge, readers may engage in the same written text in dissimilar ways. Interpretations they put on and inferences they draw from what has been read probably vary from reader to reader. Predicated on L1 reading empirical evidence that the process good readers go through differs from that poor readers do, considerable L2 learning theorists, since the 1970s, have been advocating reading strategy instruction so as to enable L2 readers to read better (Carrell, 1989). In addition, L2 reading researchers have identified reading strategies that successful L2 readers deploy and located differences in strategy use between successful and less successful L2 readers.

2.4.1 Variances in strategy use between successful and less successful L2 readers

Several verbal report studies have been conducted to provide insights into the discrepancies in reading strategy deployment between successful L2 readers and less successful L2 readers. For example, using think-aloud procedures, Hosenfeld (1977) found that successful readers approached the text in a main-meaning manner such as bearing the meaning of the passage in mind while processing L2 written texts, skipping less important words, and possessing positive self-concepts as readers. In contrast, less successful readers processed the text at the word level, lost the meaning of sentences, seldom skipped less important words, and held negative self-concepts as readers.

Similarly, Block's (1986) qualitative study showed that proficient ESL readers tended to be *Integrators*, who monitored their comprehension frequently and aggressively, integrated information, and responded to the text in an extensive mode by using the information provided by the text. By contrast, less proficient ESL readers had a tendency to be *Nonintegrators*, who banked a lot on their personal experiences to assist in their understanding the text. It is worth noting that Block's study gives evidence for L2 readers' monitoring their comprehension in the L2 reading course.

Block (1992) further found different monitoring patterns across proficient readers and less-proficient readers when they encountered lexical problems. Proficient ESL readers were inclined to draw upon syntactic clues as well as background knowledge, reread the sentence and tried to make sense of the words from context. They appeared to deal with the problem in a way as an interactive model of reading suggests. On the contrary, most of the less proficient ESL readers made little effort to work out the meanings of words, just committing themselves to identifying lexical problems. Block's work is important, since it implies that readers' insufficient L2 knowledge may prevent them from deploying some strategies in the L2 reading process.

In addition to verbal reports, L2 reading strategy researchers implement questionnaires to examine good readers' and poor readers' strategy employment in L2 reading. Yiğter, Sariçoban and Gürses's (2005) quantitative study indicated that overall, good readers, compared with poor readers, had more tendencies to predict the content of the text, figure out the author's purpose to make sense of the overall meaning of the text, actively interact with the clues emerging from the written text to understand its meaning, and summarize and comment what was read. The finding generally agrees with those found in the aforementioned qualitative-oriented studies in which verbal reports are utilized.

Other data collection technique was also adopted to shed light on discrepancies between competent L2 readers and less competent L2 readers in their strategy use during the L2 reading process. With the use of the cloze procedure, Hauptman (1979) found that the advanced L2 readers appeared to be more willing to employ the strategy: taking a chance to solve semantic problems. In contrast, the less proficient L2 readers tended to be more reluctant to take chances, be less able to capitalize on global textual information, and pay little attention to local cues. Since the reading process elicited by the cloze procedure differs from that in a regular reading setting to a certain degree, the findings may not be generalized to regular reading completely.

The aforementioned studies suggest that successful readers read differently from less successful readers to a certain extent. More specifically, successful L2 readers tend to make sense of the text in a global or interactive manner distinct from less successful L2 readers inclined to process the text in a local fashion, the finding which echoes that described in Baker and Brown's (1984) report about L1 reading. With the notion that good readers read differently from poor readers, Hosenfeld (1984) in her report listed good readers' strategies identified through think-aloud and introspective/retrospective approaches. Examples of these strategies are "identifying the grammatical category of words", "keeping the meaning of the passage in mind" and "read in broad phrases" (p. 233). The report also depicted a study on two less successful L2 readers, who translated word-by-word, hinged heavily upon the glossary for the meanings of unknown words, tackled the meanings of words without consulting context and failed to evaluate their guesses, when encountering unknown words. Hosenfeld instructed these readers in reading strategies, and their reading strategy deployment improved. What makes Hosenfeld's report significant is that a possibility – teaching poor L2 readers good readers' strategies to improve their strategy use in L2 reading – is demonstrated.

On the other hand, some L2 reading researchers (e.g., Cohen, 1986; 1998b; Sarig, 1987) challenged the conventional dichotomy of good and poor readers' strategies. Sarig (1987) found that coherence-detecting moves accounted most for both overall success and failure in reading tasks, implying that this form of moves led to success or failure in the completion of the task. Coherence-detecting moves involved global strategies such as deploying the prior content schemata or relying on textual schemata. It followed that the deployment of global strategies (good readers' strategies) did not necessarily bring about success in L2 reading. Interestingly, technical-aid moves (e.g., skipping, scanning, marking) were found to be more comprehension-deterring-oriented than comprehension-promoting-oriented. Sarig claimed that the conventionally good readers' (global) strategies might not necessarily result in success in the completion of a task and that reading was individual-oriented because each reader might read in a diverse manner and deploy different combinations of strategies. A key contribution that Sarig's work makes is that strategy types are not connected with L2 reading competence in a simple manner.

In summary, while the above studies give valuable insights into how successful and less successful L2 readers process the text, limited information is manifested on the size of effects that strategy use has on reading task performance and on how strategy use interacts with linguistic knowledge to affect reading task performance. The dichotomous classification of reading strategies in early reading strategy research is influenced by the concepts of reading processes (i.e., the bottom-up approach and the top-down approach) detailed in Section 2.2 and linked with L2 reading proficiency. The repertoire of reading strategies that successful L2 readers employ broadly contains reading for meaning, making an inference, scanning, skimming, skipping unknown words, reading in a critical manner, guessing in a context, recognizing the structure of text, activating adequate background knowledge and monitoring comprehension. Qualitative analysis results display that successful L2 readers show an inclination to deploy these strategies, whereas quantitative analysis outcomes indicate that they tend to utilize more these strategies. It follows that there seems to be "good" and "poor" reading strategies. Then, instructing less successful readers in "good" reading strategies or training them to deploy these strategies more frequently will lead to their reading L2 texts better. The case, however, is not that simple. As illustrated in Sarig's (1987) study, the deployment of strategies does not always correspond to performance on L2 reading tasks being promoted.

2.4.2 Metacognitive awareness and L2 reading

How many strategies language learners deploy or how frequently they employ strategies does not necessarily guarantee better performance on or success in a given language task (Cohen, 1998b). In the L2 reading context, as Carrell (1992) suggests, "use of certain reading strategies does not always lead to the successful reading comprehension…" (p. 168). Something else is involved in L2 readers' strategy use as they process L2 written texts. Further, Cohen (1986) points out that:

> strategies may not be inherently good or bad for a given reader. Rather, they may or may not promote successful comprehension of a text, depending on the particular reader, the particular text, the context in which the reading is going on, and the choice of other strategies in conjunction with the chosen one (pp. 132-133).

It appears that all strategies feature their own values. Whether strategies contribute to reading task performance rests on whether they can be tapped into properly and flexibly in different settings. In order to deploy various reading strategies appropriately as well as effectively in a diversity of contexts, L2 readers need a form of capability. Such capability is referred to as *metacognitive awareness* or *metacognition of reading strategies* (Yang, 2006).

Metacognition "is the ability to make your thinking visible. It is the ability to reflect on what you know and do and what you do not know and do not do" (Anderson, 2005: 767). Metacognition consists of two dimensions: *knowing that* and *knowing how*. The former dimension concerns knowledge of one's cognition and about how to regulate that cognition. The latter dimension is concerned with "executive control functions, the actual process of regulating one's cognition" (Alexander, Schallert, & Hare, 1991: 320). Within the reading domain, knowing that can be referred to as readers' knowledge about their own cognitive resources and the compatibility between readers and reading contexts in which they are involved (Baker & Brown, 1984; Carrell, 1989). Knowing how can be viewed as "the self-control mechanisms they [readers] exercise when monitoring and regulating text comprehension" (Mokhtari & Reichard, 2002: 249).

In the course of reading, readers with metacognitive awareness assess themselves as readers, evaluate their own knowledge assets, understand the requirement of a given task, set a goal and plan how to approach the task. Then, they deploy cognitive strategies with other knowledge assets to process texts, during which metacognitive strategies are employed all the time to check their own comprehension, monitor strategy deployment, evaluate the appropriateness and effectiveness of strategy use, and then adjust their strategy employment if needed. Baker and Brown (1984) observe that in comparison with readers who lack metacognitive awareness of reading, readers who are aware of the nature of reading and of their own reading strategy use tend to be better readers. It is noteworthy that most of the time metacognitive awareness operates automatically and un-observably when reading for comprehension goes smoothly. Only when comprehension is blocked and problems occur will it become conscious and detectable.

Several L2 reading research works have examined the relationship between metacognitive awareness and L2 reading. Pertinent to this is Devine's (1984) study which addresses L2 readers' self-conceptualization of the reading process. She found that even ESL beginning learners possessed internalized models of reading: meaning-, word-, and sound-centered models. Meaning-centered readers showed better comprehension than word-centered readers on given tasks, whereas word-centered readers displayed better comprehension than sound-centered readers on given tasks. Devine's work leads us to understand that how L2 readers self-conceptualize the reading process is related to how they approach the written texts and further their reading comprehension.

A qualitative study carried out by Auerback and Paxton (1997) also showed the three models of reading identified in Devine's (1984) study. Moreover, the authors found that L2 readers more aware of resorting to interactive strategies comprehended the reading better, compared with those who simply focused on the sentence level of the text. This key finding reveals the close relationship among metacognitive awareness, strategy deployment and L2 reading. Related information is also available in the following studies.

Barnett's (1988) study indicated that both strategy use and perceived strategy use related to L2 reading comprehension positively. The more L2 readers perceived they utilized effective strategies, the better they employed reading strategies, and the more comprehension they obtained. It is evident that in the field of L2 reading strategies, certain attention has been given to explore the relation among metacognitive awareness, strategy employment and L2 reading comprehension.

Using metacognitive awareness questionnaires, Carrell (1989) examined L2 readers' metacognitive awareness in relation to reading strategy use. She found that for readers with high L2 proficiency, part of the top-down reading strategies was positively related to reading performance. On the other hand, there was, for readers with low L2 proficiency, a positive relationship between part of the bottom-up reading strategies and reading performance. The finding implied that L2 readers with higher L2 proficiency tended to have global perceptions of their partial reading strategy use, while L2 readers with lower L2 proficiency showed an inclination to possess local perceptions of part of their reading strategy employment. Carrell's study is significant, because it suggests that L2 proficiency may be related to L2 readers' metacognitive perceptions of strategy use which is associated with their L2 reading performance.

The effect of L2 readers' perceptions of strategy use on their reading comprehension performance was investigated in Padron and Waxman's (1988) study in which questionnaires were administered. With regression analysis, the authors found that two of the negative strategies – "thinking about something else while reading" and "saying the main idea over and over" (p.147) – were negative predicators of reading test performance. Clearly, Padron and Waxman's work supplies us with empirical evidence that L2 readers' perceptions of strategies they employ can function as a predictor of their

L2 reading comprehension performance.

Jiménez, García and Pearson's (1996) qualitative study illustrated that successful L2 readers were more inclined to deploy global reading strategies as mentioned in Section 2.3.1 and were more aware of the differences and similarities in L1 and L2. Jiménez *et al.* argued that due to this awareness, successful L2 readers might deploy more appropriate strategies and perform better on L2 reading. It follows that metacognitve awareness plays a crucial role in distinguishing successful readers from less successful readers. This notion is supported by the evidence provided by quantitative research.

In a more quantitative style study, Sheorey and Mokhtari (2001) found that both L1 readers and L2 readers with high reading ability showed the comparably higher degree of reported deployment of cognitive and metacognitve reading strategies than L2 readers with lower reading ability. Importantly, both L1 and L2 readers displayed awareness of almost all of the strategies covered by the questionnaire. Sheorey and Mokhtari's study implies that L2 reading strategy researchers have shifted their attention onto investigating the similarities and differences in metacognitive awareness of strategy use of L1 and L2 readers.

Following this line of research, Mokhtari and Reichard (2004), with the administration of the questionnaire, found that both L1 readers and L2 readers at an advanced level of L2 proficiency (equivalent to a score of 500-550 of the TOEFL test) displayed similar patterns regarding the awareness of their strategy use and reported strategy deployment. L2 readers with an advanced level of L2 proficiency appeared engaged in a strategic reading process not different from that L1 readers were engaged in. However, like Sheorey and Mokhtari's (2001) study, no task was present for participants to refer to during the strategy use elicitation procedure. Then, the gathered data could be questionable, given that it may be demanding for participants to make a decision on the extent of their strategy use.

To summarize, the studies which have been discussed thus far provides qualitative and quantitative evidence that L2 readers' metacognitive awareness relates to and may impact upon their reading strategy deployment and their reading performance. As stated at the inception of this section, strategy use is not equivalent to success in L2 reading. L2 readers not only need to possess a repertoire of strategies at their disposal but also need to have metacognitive awareness during the L2 reading process. In other words, they ought to be aware of their goals, monitor their reading process, check their reading comprehension, deploy strategies if necessary, evaluate their strategy deployment, and if needed adjust their strategy use after evaluation. Without metacognitive awareness or with less this awareness, L2 readers may deploy strategies inappropriately or ineffectively. They perhaps fail to overcome obstacles to their comprehension in L2 reading even though they deploy strategies. Metacognitve awareness also results in L2 readers' strategy deployment being more flexible. Being flexible about strategy use denotes that strategy deployment not merely varies with tasks, but combines with other strategy use appropriately. This flexibility of strategy employment may increase the chance of success in L2 reading. In this regard, Yang's (2006) qualitative study provides relevant evidence.

Utilizing think-aloud and retrospective verbal reports, Yang (2006) found that only when L2 readers employed strategies on specific occasions did their deployment of strategies contribute to L2 reading. Otherwise, despite their employment of some reading strategies, L2 readers might still fail to make sense of the text. This suggests that strategy deployment warrants being duly checked and assessed with the activation of metacognitive awareness as mentioned above. Another important finding by Yang was that L2 readers occasionally employed both reading and comprehension monitoring strategies simultaneously in the reading process to detect and resolve reading problems they encountered. The finding echoes what Macaro (2004; 2006), and Schraw and Moshman (1995) remark – metacognitive strategies (e.g., monitoring or evaluating), with a view to orchestrating cognitive activities, are often included in a strategy group which consists of cognitive strategies.

Apart from differences in strategy deployment between successful L2 readers and less successful L2 readers and the magnitude of metacognitive awareness in L2 reading, L2 reading strategy researchers have also shown interest in exploring the relation of strategy employment to readers' other attributes, particularly to L2 language knowledge or L2 proficiency. Prior to discussing the relationship between L2 language knowledge and strategy employment in the L2 reading context, the role of L2 language knowledge or L2 proficiency, particularly lexical knowledge and grammatical knowledge, in L2 reading is reviewed, given that in the current study language knowledge is limited to lexical knowledge and grammatical knowledge.

2.5 L2 Language knowledge, particularly lexical knowledge and grammatical knowledge and L2 reading

Language knowledge refers to "a domain of information in memory that is available for use…in creating and interpreting discourse in language use" (Bachman &

Palmer, 1996: 67). It is conceivable "that, if readers do not know the language of the text, then they will have great difficulty in processing the text" (Alderson, 2000: 34). "The ease with which the language of a particular text can be processed…depend[s] upon the nature of the reader's linguistic knowledge" (ibid.). Some L1 reading process models such as Stanovich's interactive-compensatory reading model (1980) or Rumelhart's (1977) interactive model of reading imply that reading comprehension is greatly likely to be impeded if readers lack sufficient language knowledge (e.g., lexical knowledge or grammatical knowledge) to process written texts efficiently. Influenced by L1 reading process models like those mentioned above and by the fact that more and more EFL or ESL readers need to read specialized texts in English, the early L2 reading studies have examined L2 readers' reading problems and thereby demonstrated the importance of language knowledge, particularly lexical knowledge and grammatical knowledge in L2 reading.

In a study where questionnaires were used, Yorio (1971) found that in comparison with grammar, vocabulary was more challenging for L2 readers in their L2 reading. Yorio explained that L2 readers could acquire most of syntactic knowledge of an L2 and even master it through persistent learning, because grammatical knowledge was more systematic and finite. However, it was quite difficult for L2 readers to master lexical knowledge due to its less systematic and infinite nature.

On the other hand, through interviews, Cohen, Glasman, Rosenbaum-Cohen, Ferara, and Fine (1979) found that L2 readers with advanced L2 proficiency, when they read specialized texts, often failed to pick up on conjunctive words. With such a finding, Cohen *et al.* argued that L2 readers read more locally than L1 readers and they had trouble synthesizing information across sentences and paragraphs. They also uncovered that these L2 readers had difficulty in processing sentences containing noun clauses and decoding the meanings of nontechnical words in the technical texts they read. Cohen *et al.*'s findings highlight the weight that grammatical knowledge carries in L2 reading.

In addition to identifying L2 reading problems relevant to vocabulary and grammar, attention has been given to the relationship between lexical knowledge and grammatical knowledge in L2 reading. Barnett's (1986) study showed that reading recall ability varied, along with the level of vocabulary knowledge and of syntactic knowledge in L2. A contribution of Barnett study is that evidence is provided that both lexical knowledge and syntactic knowledge are linked to L2 reading comprehension and these two types of language knowledge interact with each other in L2 reading comprehension.

More valuable insights into the linkage between lexical knowledge and grammatical knowledge in L2 reading are given in Nassaji's (2003a) and Paribakht's (2004) research works. Nassaji's (ibid.) found that in L2 reading, grammatical knowledge functioned as a type of knowledge that L2 readers drew upon to infer the meanings of unknown words, although they seldom relied on this kind of knowledge and they might not succeed in figuring out the meanings of words with the use of this form of knowledge. Nassaji's (ibid.) finding is distinct from Paribakht's (ibid.) to a certain degree.

Paribakht (2004) reported that L2 readers tapped into a diversity of knowledge sources available to process the meaning of the lexis. Among these knowledge sources, sentence-level grammatical knowledge accounted for most (35%) – the grammatical knowledge was defined as "the knowledge of speech parts and syntactic relationships among words within a sentence" (p. 152). Paribakht concluded that L2 readers' grammatical knowledge might have an impact on L2 lexical inferencing processing in L2 reading and also contribute to the utilization of L2 readers' strategic competence.

With the application of structural equation modeling, Shiotsu and Weir (2007) found that knowledge of syntax significantly contributed more to L2 reading test performance than knowledge of vocabulary. This finding, combined with the findings in Nassaji's (2003a) and Paribakht's (2004) studies, suggests that grammatical knowledge carries more weight than lexical language in L2 reading. However, the following studies show a different story.

Based on a two-year longitudinal study, Droop and Verhoeven (2003) found that L2 readers' previous vocabulary knowledge had slightly more impacts on reading comprehension performance than previous morphosyntactic knowledge at the end of the elementary school third grade. In addition, their previous vocabulary knowledge yielded more effects on reading comprehension performance than previous morphosyntactic knowledge at the end of the fourth grade. These findings can be taken as an indication that for L2 readers, vocabulary knowledge plays a more pivotal role in L2 reading comprehension than grammatical knowledge.

Nassaji's (2003b) quantitative study showed that lexical-semantic knowledge functioned most in distinguishing skilled from less skilled readers, followed by the word-recognition skill. The finding suggests that L2 lexical knowledge bears more importance in L2 reading than other types of language knowledge, which accords with Droop and Verhoeven's finding (2003), but goes against what Nassaji's (2003a), Paribakht's (2004) and Shiotsu and Weir's (2007) studies have implied or indicated. Clearly, the relative importance of lexical knowledge and grammatical knowledge in L2 reading is still inconclusive.

L2 reading literature has also manifested what a language threshold is for L2 academic reading. Laufer and Sim (1985) found that for EFL readers, the language threshold for reading academic texts in English was reflected in a 65-70% score on the reading comprehension section of the FCE exam (the First Certificate of English Exam). Another vital finding was that the construct of the language threshold for reading academic texts consisted of vocabulary, knowledge of the subject matter, discourse markers and syntactic structure. This finding highlights the weight that L2 lexical knowledge and L2 syntactic knowledge give to L2 reading.

To summarize, most of the aforementioned studies do not provide empirical evidence for the notion that lexical knowledge and grammatical knowledge have an effect on L2 reading in a way that the effect is shown in a single modeling framework. Nevertheless, they still evidence this notion through other data analysis methods such as multiple regression analysis or analysis of variance. Further, the studies discussed above show crucial roles of L2 lexical knowledge and L2 grammatical knowledge in L2 reading, the relationship between them in L2 reading, and their contributions to L2 reading performance. An attempt has been made to explore the relative significance between lexical knowledge and grammatical knowledge in L2 reading and that between these types of language knowledge and other components that affect L2 reading. It is not challenged that L2 readers must be equipped with a certain level of L2 language knowledge in order to perform L2 reading successfully and smoothly (Devine, 1988). L2 lexical knowledge and grammatical knowledge, parts of language knowledge, are key language assets that L2 readers access to make sense of L2 texts. In the following section, the relation between L2 proficiency/L2 language knowledge and L2 readers' strategy use will be addressed.

2.6 L2 proficiency/L2 language knowledge and strategy use in L2 reading

L2 Strategy research has demonstrated that strategy deployment is related to L2 proficiency/L2 language knowledge (e.g., Bialystok, 1981; Green & Oxford, 1995; Griffiths, 2003; Hong-Nam & Leavell, 2006). However, their relationship is not always positive. A good illustration for this is a pattern of strategy use among advanced language learners, intermediate language learners and beginning language learners, identified by Green (1991, cited in Oxford & Cohen, 1992).

> Green reports that his advanced language learners often have significantly lower strategy use than intermediate language learners, and that intermediates use strategies significantly more than do beginners. Thus, strategy use in Green's study might appear to be curvilinear, with intermediates using language learning strategies far more than advanced and beginning language learners. One might speculate that advanced learners might have automatized their learning behaviors, so they might not use or need language learning strategies as much as do intermediates; and beginners might not have yet developed a large, conscious, and frequently tapped repertoire of strategies. (Oxford & Cohen, 1992: 13)

This identified pattern indicates that strategy deployment varies with language learners at different language levels. In terms of the concept of a language threshold, it appears that there are two language thresholds for strategy use among these language learners. One (a lower language threshold) exists between intermediate language learners and beginning language learners, whereas the other (an upper language threshold) is present between advanced language learners and intermediate language learners. Crossing a lower language threshold is a prerequisite for language learners to tap into strategies, while reported strategy use is decreasing once learners cross an upper language threshold. On the other hand, from an information-processing perspective, it can be argued that advanced language learners might deploy their strategies on an automatic-process basis. Thus, they report less strategy use than intermediate language learners. By contrast, intermediate language learners could capitalize on their strategies on a controlled-process basis; consequently, they report more strategy use compared with advanced language learners.

It follows that within the L2 context, L2 learners' L2 proficiency makes a difference to their strategy use. In the initial phase of the development of L2 proficiency, L2 learners utilize strategies comparatively consciously but limitedly, less efficiently and less sophisticatedly in their L2 learning and L2 use. Their strategy deployment is not bound to contribute to performing tasks or solving problems they encounter. This is because their limited L2 proficiency presents them from successfully accessing their strategy assets accumulated during the L1 learning process or from appropriately and effectually drawing upon strategies developed in the L1 learning course in the L2 setting. When L2 learners' language proficiency betters and their strategy deployment advances from a stage of nonmastery to mastery in the L2 context, "their strategic processing becomes skillful processing (i.e., unconscious competence)" (Phakiti, 2003: 670). Then, strategy employment, for them, taxes less cognitive capacity. Their strategy deployment turns more flexible, efficient, and effective. Once learners' strategy use becomes automatic or their reliance on strategies gets less frequent, it will become difficult to detect their strategy use. However, for competent L2 learners, strategic processing still carries significance. Herculean tasks that override L2 learners' current language proficiency will drive them to reactivate strategic awareness and capitalize on strategies to deal with the tasks.

In the L2 reading context, L2 readers' reading strategies expand with their L2 proficiency/L2 language knowledge. Competent L2 readers, with the progress of their L2 proficiency/L2 language knowledge, develop well in their strategy use and possess their own repertoires of reading strategies. When they process a less demanding written text, they tend to deploy and report fewer reading strategies. This is due to the fact that they can rely on other cognitive resources such as L2 language knowledge to make sense of the text without difficulty, or they employ some reading strategies in an automatized way. Their partial reading strategies are unconscious and undetectable. Only when they encounter cognitive challenging tasks will the reading strategies turn conscious and observable. As for less competent L2 readers, reading strategy employment has not well-developed yet in L2 reading. Accordingly, their strategy use is limited and more conscious, and easy to observe. On account of their limited L2 proficiency/L2 language knowledge, they often focus on utilizing local reading strategies to tackle unknown words. The following studies exemplify variances in strategy deployment among L2 readers with different L2 proficiency.

Cziko (1980) reported that L2 readers at an intermediate level of L2 proficiency, less aware of contextual information, tended to employ a bottom-up approach of reading with heavy dependence on graphic information. In contrast, readers at an advanced level of L2 proficiency, sensitive to contextual information, were inclined to draw on an interactive approach of reading with reliance upon contextual and graphic information. Cziko's work highlights differences in how L2 texts are processed among L2 readers with discrepant L2 proficiency.

McLeod and McLaughlin (1986) found that beginning L2 learners focused on local information and made few predictions, so that they made more errors in a cloze test. On the contrary, advanced L2 learners tended to make predictions in the reading process.

The aforementioned two studies show that in the field of L2 reading, an attempt has been made to uncover reading behaviors of L2 readers with different L2 proficiency in L2 reading. The following studies provide more information for variations in strategy use in L2 reading among L2 readers with differential L2 proficiency.

Stevenson, Schoonen and Glopper (2003) published the finding that regardless of their L2 reading ability, L2 readers were capable of capitalizing on Regulatory (i.e., metacognitve) strategies in L2 reading, particularly monitoring strategies. However, rather than content-oriented ones which readers used to construct the global representation of the text, most these Regulatory strategies were language-oriented ones that readers deployed to process linguistic components and relations in the text. The authors explained that insufficient L2 proficiency of the study participants was responsible for the language-oriented Regulatory strategies being largely deployed in L2 reading. Stevenson *et al.*'s work implies that L2 proficiency is linked to L2 readers' metacognitive strategy use.

Based on thin-aloud data, Davis and Bistodeau (1993) found that readers with less L2 proficiency utilized more bottom-up strategies and fewer top-down strategies in their L2 reading. Compared with Stevenson *et al.*'s (2003), Davis and Bistodeau's study shows direct evidence for the notion that a relationship is present between L2 learners' L2 proficiency and part of their strategy deployment in L2 reading.

In a study where think-aloud protocols and retrospective interviews were used, Upton and Lee-Thompson (2001) found that the beneficial deployment of strategies enhanced with the growth of L2 proficiency. Although compared with the intermediate ESL group and the advanced ESL group, the post-ESL group was scarcely inclined to depend on L1 as a reading strategy and reported fewer reading strategies, they employed them relatively effectively and beneficially once they tap into them. The finding lends support to what has been mentioned at the beginning of this section – the development of strategy use in the L2 context.

The following studies where questionnaires are implemented to gather strategy data also offer related information to support the notion that L2 proficiency is associated with strategy use in L2 reading. Oxford, Cho, Leung, and Kim's (2004) work revealed three differences between high L2 proficiency readers and low L2 proficiency ones in their strategy use. Firstly, overall L2 readers with low L2 proficiency reported higher frequency of strategy deployment on most strategy use items than those with high L2 proficiency did. Secondly, L2 readers with high L2 proficiency reported deploying more often on top-down and metacognition-oriented strategy items, while those with low L2 proficiency on bottom-up strategy items. Finally, L2 readers with high L2 proficiency were significantly distinct from those with low L2 proficiency in the frequency of overall strategy use in the Difficult Task condition – low-proficiency readers reported employing strategies more frequently than high-proficiency readers.

Different from Oxford *et al.* (2004), Sheorey and Edit (2004) found that EFL readers with high English proficiency self-reported the higher frequency of their strategy use than those with low English proficiency on two-thirds of strategy use items as they processed EFL academic texts. A statistically significant difference was identified between the high English proficiency group and the low English proficiency group in the strategy use frequency on one-third of reading strategy use items and each of the three strategy use categories (i.e., global reading strategies; support strategies; and problem-solving strategies). Sheorey and Edit's study contributes to our understanding that there is a positive and reciprocal relation among reading ability, language proficiency and reading strategy awareness.

In summary, the studies discussed above give limited direct evidence for whether L2 language knowledge has an influence on strategy employment in L2 reading and vice versa. Further, they fail to offer an overall picture of the linkage amongst L2 language knowledge, strategy use and L2 reading performance for groups with different L2 proficiency/ability. Nonetheless, these studies suggest that L2 proficiency carries weight in L2 reading strategy deployment. Although sometimes strategy deployment is not positively correlated with L2 proficiency, for L2 readers, crossing a certain level of L2 proficiency or possessing a certain amount of L2 language knowledge appears a precondition for deploying reading strategies appropriately. This is related to a language threshold for transferring L1 reading ability/strategies to L2 reading, which is addressed as follows.

2.7 A language threshold for transferring L1 reading ability/strategies to L2 reading

The focus of substantial discussion in the past three decades is whether reading in L2/FL is a reading problem or a language problem. Related to this focus is a *language threshold* or a *language competence ceiling* for transferring L1 reading ability to L2 reading. The *language threshold* hypothesizes that L2 readers need to have sufficient L2 proficiency/L2 language knowledge in order to apply L1 reading ability/strategies appropriately to their L2 reading (Grabe & Stoller, 2002). Clarke, in 1979, first extended the notion of a language threshold to L2/FL reading – limited L2 proficiency short-circuits the transfer of reading ability acquired in L1 to L2. Clarke (1979; 1980) found that although good L1 readers performed better than poor L1 readers on L1 and L2 reading, good L1 readers' advantage over poor L1 readers diminished on L2 text performance compared with that on L1 text performance. He concluded that the existence of a "language competence ceiling" (1979: 138) in L2 prevented good L1 readers from taking advantage of "effective reading behaviors in the target language" (ibid.). More specifically, deficient L2 proficiency short-circuited the reading mechanism of good L1 readers, which forced them to invoke poor reading strategies utilized in L1 reading when they tackled more challenging tasks in L2 reading. Coady (1979) and Hauptman (1979) conducted similar research on L1-L2 reading; however, they failed to concretely point out the concept of a language threshold for transferring L1 reading ability/strategies to L2 reading.

Alderson's (1984) seminal article addressing the focus stated above draws plenty of L2/FL reading researchers' attention. After reviewing several studies, he concluded that crossing a certain level of L2 was a prerequisite prior to transferring L1 reading ability to L2 reading – reading in L2 was a language problem. Later, in the field of L2/FL reading, a substantial amount of research has been undertaken to explore the relationship among L1 reading ability, L2 proficiency/L2 language knowledge and L2 reading performance. For example, Perkins, Brutten and Pohlmann (1989) carried out a study to investigate the relationship between L1 reading and L2 reading. They found that the correlation between L1 reading and L2 reading at the low L2 proficiency level was lower than that at the middle level and that at the middle level was lower than that at the high level. They also found that the correlation between L1 reading and L2 reading at the low L2 proficiency level did not reach statistical significance, but that at the middle level and at the high level did. They concluded that it was at the high L2 proficiency level where the transfer of L1 reading ability to L2 reading occurred. Their conclusion, however, is not convincing because, strictly speaking, their study just displays the relation between L1 reading ability and L2 reading performance at three different L2 proficiency levels.

Carrell (1991), through a multiple regression analysis, found that both L1 reading ability and L2 proficiency level served as significant predictors of L2 reading ability. She also found that L1 reading ability functioned as a stronger predictor of L2 reading than L2 proficiency in the L2 group, while the case reversed in the FL group. Carrell attributed this variation to the fact that the L2 group lived in a "second language environment" (an English as a second language environment), while the FL group lived in a "foreign language environment" (a Spanish as a foreign language environment). A language threshold in the L2 setting was lower than that within the FL context so that L2 learners were able to arrive at and cross it with less effort. Then, L1 reading ability could be transferred to L2 reading in the L2 context. Carrell's study highlights that a target language learning environment probably impacts on a language threshold for transferring L1 reading ability to target language reading. However, due to participants with different L1 backgrounds, it is difficult to make a within-participant comparison, which limits the finding of the study.

To overcome the drawback of Carrell's study, Bossers (1991) undertook a study in which participants had similar L1 background and their L2 language knowledge was measured by a standardized test. He found that both L1 reading ability and L2 language knowledge played a significant part in L2 reading and L2 language knowledge acted as a more powerful predicator of L2 reading than did L1 reading ability. The latter finding contradicts Carrell's (1991). This difference may be due to variances in participants' L2 proficiency between these two studies. In addition, he found that at the initial stage of L2 development, the importance of L2 language knowledge outweighed that of L1 reading ability in L2 reading; however, at the advanced stage of L2 development, L1 reading ability bore more importance than L2 language knowledge. Bossers concluded that in the advanced phase of L2 development, L2 readers crossed a language threshold and transferred L1 reading ability to L2 reading.

Consistent with Bossers's (1991) and Carrell's (1991) studies, the finding that both L1 reading ability and L2 proficiency functioned as significant contributors to L2 reading performance was given in Bernhardt and Kamil (1995) study. Another important finding was that L2 proficiency, compared with L1 reading ability, performed as a stronger predictor of L2 reading comprehension. Like Alderson (1984), Bernhardt and Kamil claimed that reading in L2 was a language threshold problem.

After analyzing the data collected from junior high school students, Lee and Schallert's (1997) came to a conclusion that L2 readers needed to accumulate L2 language knowledge to a certain extent before they were able to successfully capitalize on their L1 reading ability to assist in their L2 reading comprehension. However, their conclusion appears dubious. This is because their conclusion is based on the evidence that the relationship between L1 reading ability and L2 reading performance became slightly higher and higher along with the increase of L2 proficiency. This simply suggests a simple linear relationship between L1 reading ability and L2 reading performance at different L2 proficiency levels, just as Perkins *et al.*'s (1989) study has indicated.

More relevant information is provided by Pichette, Segalowitz and Connors' (2003) longitudinal study. From the first test result, they found that neither L1 reading ability nor L2 language knowledge functioned as a significant contributor to L2 reading performance for the high L2 language knowledge group, whereas L2 language knowledge served as a significant contributor to L2 reading performance for the low L2 language knowledge group. Interestingly, from the second test result, they found that L1 reading ability contributed significantly to L2 reading performance for the high L2 language knowledge group, while L2 language knowledge still significantly contributed to L2 reading performance for the low L2 language knowledge group. Pichette *et al.* concluded that L2 readers' failure in possessing a sufficient amount of L2 language knowledge appeared to short-circuit their transfer of L1 reading ability to L2 reading.

Whether a language threshold for transferring L1 reading ability to L2 reading varies from a task to another task is also investigated. Taillefer (1996) found that both L1 reading ability and L2 proficiency acted as predictors of L2 reading performance but to a divergent extent in different tasks. For readers at a high level of L2 proficiency, L1 reading ability contributed more to performance on an easy task (operationalized by an L2 scanning test), while L2 proficiency contributed more to performance on a difficult task (operationalized by a receptive reading test). However, for readers at a low level of L2 proficiency, neither L1 reading ability nor L2 proficiency served as a significant predictor of performance on an easy or a difficult task. Taillefer's study gives empirical evidence for the concept that a language threshold for tapping into L1 reading ability in L2 reading is subject to task difficulty and readers' L2 proficiency.

From a cognitive psychology perspective, Walter (2004) reported that some L2 readers comprehended L2 texts poorly, even though they were equipped with reading comprehension skills and able to comprehend L1 texts well. She explained that the reason why some L2 readers were capable of constructing mental representations of texts in L1 but not in L2 consisted in that their L2 proficiency failed to attain a certain level at which reading comprehension skills employed well in L1 reading can be accessed and applied to L2 reading. Walter's work lends additional support to the existence of a language threshold for transferring L1 reading ability to L2 reading.

In addition to measuring participants' L1 reading ability by L1 reading tests, Yamashita (2002) collected and analyzed participants' strategies used in L1 and L2 reading, which was distinct from the aforementioned research works. She found that L1 reading ability had a positive impact on L2 reading performance, but the effect was less strong than that of L2 proficiency. This finding is consistent with that in previous studies (e.g., Bernhardt & Kamil, 1995; Bossers, 1991; Lee & Schallert, 1997). Another finding was that readers with low L2 proficiency failed to take advantage of promoting strategies (i.e., strategies contribute to correct comprehension) in L2 reading even though they had high L1 reading ability. With this finding, she concluded that there was a language threshold for deploying strategies appropriately. Substantively, Yamashita classified reading and test-taking strategies into language dependent strategies and language independent strategies based on whether these strategies require linguistic processing – she failed to remark what linguistic processing is required. She submitted that the extent to which language dependent strategies were transferred to L2 reading relied on L2 readers' L2 proficiency. In contrast, language independent strategies appeared to be independent of L2 readers' L2 proficiency. Yamashita' study sheds light on the notion that a language threshold is just closely related to partial strategy use, not all strategy use in L2 reading, as Davis and Bistodeau's (1993) and Stevenson *et al.*'s (2003) studies imply.

In summary, a clear picture has been gained that both L1 reading ability/strategies and L2 proficiency/L2 language knowledge come into play in L2 reading performance and L2 proficiency/L2 language knowledge usually contributes more to L2 reading performance. Further, in L2 reading, prior to drawing upon L1 reading ability/strategies, L2 readers need to cross a language threshold of L2 proficiency. Limited L2 proficiency will obstruct them from transferring L1 reading ability/strategies to L2 reading. Additionally, a language threshold is regarded as a relative, not static form, which is subject to move according to readers' motivation, background knowledge, purposes (Hudson, 1982; Kern, 1989), or the nature of given tasks. When the concept of a language threshold is applied to reading strategy deployment, it follows that some reading strategy use in L2 reading relies on certain amount of L2 language knowledge. However, in most of the relevant studies, what is gathered is related to L1 reading ability assessed by an L1 reading test. Therefore, collected data has nothing to do strategy deployment. Because of this shortcoming, these studies fail to provide appropriate empirical evidence for the issue – whether there is a language threshold for some reading strategy employment to contribute to L2 reading performance.

2.8 Effects of test methods on L2 reading comprehension

Reading comprehension has been measured by a variety of test formats such as cloze tests, multiple-choice tests or written recall tests. As Alderson (2000) suggests, there is no best technique to assess reading comprehension. Reading comprehension measured by a test format such as a cloze test varies to some extent from that gauged by another test format such as a multiple-choice test, given that test-takers' reactions induced by divergent test methods are not totally equivalent. Bachman (1990) observes that differences in language test performance are attributed to variations in the characteristics of test tasks or test methods, in addition to test-takers' characteristics. His remarks indicate that in the domain of language testing, test formats have been recognized to have an effect on how well test-takers perform a test. Given that the current study investigates students' strategy use in a multiple-choice reading test, it is necessary to review literature germane to test-takers' strategy use in a multiple-choice reading test. However, prior to that, it is worthwhile to pause to discuss components which impact upon test performance, seeing that this study explore the linkage among students' language knowledge, strategy use and their performance on a reading test.

In the last four decades, language assessment researchers have examined the relationships among cognitive variables, language use and given tasks to explore factors that influence language test performance and to depict the nature of language proficiency (Purpura, 1997; 1999). The general consensus has been reached that test constructors' assumptions about what their tests measure are not completely equal to what their tests really assess. Or their expectations of how test-takers respond to test items do not fully correspond to how test-takers actually sit tests. Bachman (1990) has described several factors that influence test scores, which supplies us with a proper understanding that test scores represent and involve more than what tests are purported to assess. In addition, Bachman and Palmer's (1996) model of language ability in language use and language test performance has been put forward to address conceptual components operating in a test-taking situation. Given an attempt to examine the relationship among students' English language knowledge, strategy deployment and their EFL multiple-choice reading test performance, the model and the aforementioned Bachman's (1990) factors which affect test scores are adopted as the other part of the theoretical frameworks of this study and addressed as follows.

2.8.1 Bachman's factors that affect test scores

There has been recognition that test results contain not merely what tests are purported to gauge but something else as well. In order to supply test constructors or language teachers with a proper understanding of what tests they design measure, Bachman (1990) profiles four types of factors that exert an effect on test results: *communicative language ability*, *test method factors*, *personal attributes* and *random* *factors*. *Communicative language ability* consists of language competence (i.e., language knowledge in Bachman and Palmer's (1996) model), strategic competence and psychophysiological mechanisms. *Test method factors* concerns the characteristics of the test tasks functioning to elicit test performance. *Personal attributes* comprise culture, attitudes, cognitive style, strategy use and such like. Finally, *random factors* relate to unpredictable events occurring during a test, test-takers' physical or mental conditions during a test, or measurement error.

Several studies have been conducted, grounded on Bachman's (1990) four categories of factors that impact on test scores (e.g., Kunnan, 1995; Nieh, 2003; Purpura, 1997; 1999). In the current study, factors regarding communicative language ability and personal attributes were concerned with. More specifically, within communicative language ability, language knowledge and strategic competence were focused on, whereas, in the personal attributes, strategy employment was centered on – all of these were discussed briefly in Sections 2.8.2.1, 2.8.2.2 and 2.8.2.3. The following figure characterizes the impacts of these four forms of factors on test scores. (What the current study focuses on is boldfaced.)

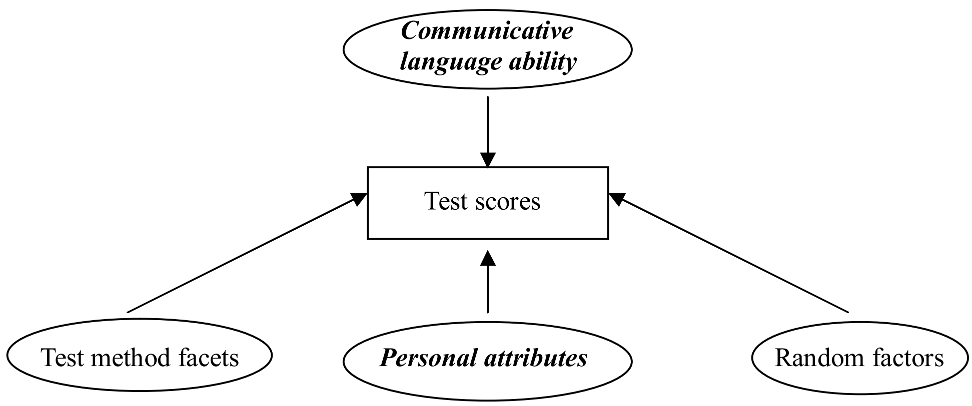


Figure 2.1 Bachman's (1990) factors that influence test scores

According to Bachman (1990), ovals referred to as observed variables represent four types of factors that impact upon test performance. A rectangle signifying an unobserved variable is concerned with test scores. Single-headed arrows symbolize hypothesized causal relationship between factors and test results. Based on the above diagram, in the present study, English language knowledge and strategy deployment were postulated to have an effect on reading test performance.

2.8.2 Bachman and Palmer's model of language ability in language use and language test performance

Bachman and Palmer (1996) provide a model which depicts the relationship among factors that affect language use and language test performance. This revised model is grounded on Bachman's (1990) model of communicative language ability. A number of research works related to L2 assessment or strategy deployment in a test context have been predicated on this model (e.g., In'nami, 2006; Kobayashi, 2002; Nikolov, 2006; Phakiti, 2003). Figure 2.2 shows Bachman and Palmer's (1996) model of language ability in language use and language test performance. (What the current study focuses on is boldfaced.)

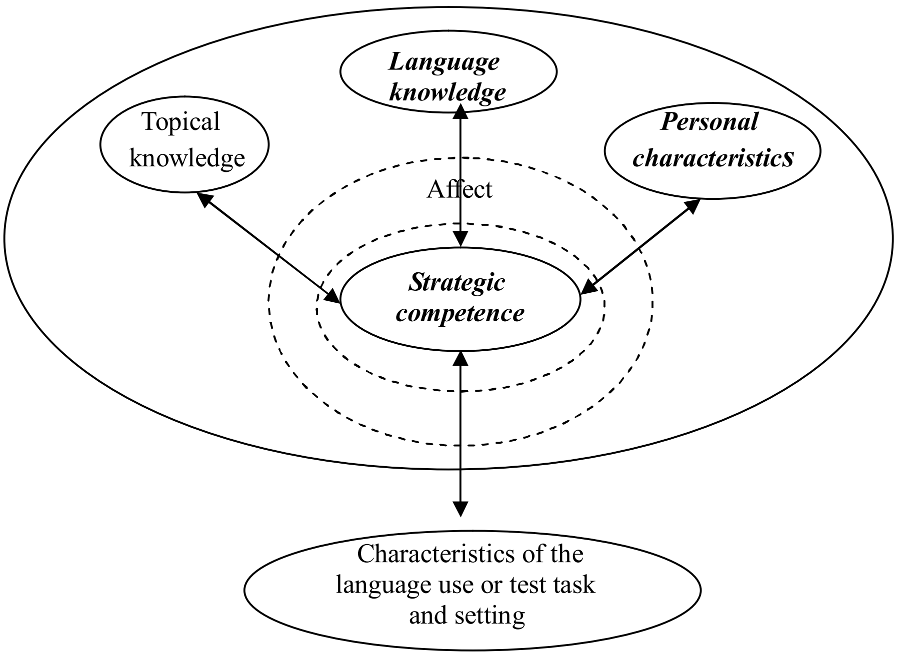


Figure 2.2 Bachman and Palmer's (1996) model of language ability in language use and language test performance

As illustrated in Figure 2.2, the largest oval represents test-taker based components that influence test performance. It embodies four small ovals and two broken line ovals, which correspond to individual test-taker based components subsuming topical knowledge, language knowledge, strategic competence, personal characteristics, and affect (i.e., emotional factors). Distinct from Bachman's (1990) model of language ability in which language knowledge exerts an influence on strategic competence, Bachman and Palmer's (1996) model indicates that language knowledge yields an effect on strategic competence and vice versa. In addition to language knowledge, topical knowledge and personal characteristics respectively interact with strategic competence. These interactive relations are symbolized by bidirectional arrows shown in Figure 2.2. Below the largest oval is a middle oval denoting the characteristics of language use, test tasks or test contexts, which interact with strategic competence. Such a relationship is also represented by a bidirectional arrow. Finally, affect within the largest oval, drawn surrounding strategic competence with broken line ovals, moderates the relationships of strategic competence to (a) topical knowledge, (b) language knowledge, (c) personal characteristics, and (d) the characteristics of language use, test tasks or test contexts.

Bachman and Palmer's (1996) model gives a clear picture of what components are involved in language test performance as well as how these components interact with each other. Their model can be utilized as a checklist to assist test constructors and language teachers in designing and developing language tests, as Bachman and Palmer (ibid.) suggest (for details, see Bachman and Palmer, 1996: 76-77). However, their model features a limitation which is related to the construct of strategic competence. Within their model, strategic competence is defined as metacognitive strategies, solely involving a set of metacognitive components (i.e., planning, assessment and goal-setting). Both Skehan (1998) and Purpura (1999) have pointed out that defining strategic competence as an array of metacognitive strategies has its drawback in that it is not grounded on empirical research works. Several L2 language testing researchers have also called for amending the construct of strategic competence based on the findings of their studies (e.g., Nikolov, 2006; Phakiti, 2003; Purpura, 1997; 1999). In the current study, a suggestion was provided for the construct of strategic competence (see Section 6.2).

The present study investigated the relationship among test-taker based factors, focusing on language knowledge, strategic competence and personal characteristics (more specifically, strategy deployment), and reading test performance. Language knowledge, strategic competence and personal characteristics were operationalized by an English language knowledge test and a reading and test-taking strategy use questionnaire, whilst reading test performance was operationalized by a multiple-choice reading comprehension test. In the following sections, language knowledge, strategic competence, and personal characteristics centering on strategy deployment will further be discussed.

2.8.2.1 Language knowledge

Within Bachman and Palmer's (1996) model of language ability, *language knowledge* is referred to "as a domain of information in memory that is available for use by the metacognitive strategies in creating and interpreting discourse in language use" (Bachman & Palmer, 1996: 67). Language knowledge consists of organizational knowledge and pragmatic knowledge. Organizational knowledge comprises grammatical knowledge and textual knowledge. It is concerned with formal structures of languages, by which to generate or understand grammatically acceptable utterances or sentences, as well as to organize these to construct the meaning of what has been processed or what is going to be expressed. On the other hand, pragmatic knowledge encompasses functional knowledge and socialinguistic knowledge. It functions to construct or interpret the meaning of what is being processed by means of connecting sentences or utterances and texts with their own meanings, with language users' intentions and with features germane to language use contexts. In a language testing situation, all components interact not merely with one another but with features of test settings or test methods as well (Bachman & Palmer, 1996).

Although Bachman and Palmer's (1996) categorization proffers a general picture of what components language knowledge encompasses, language knowledge, in fact, may not be demarcated as definitely as they describe. For example, Bachman and Palmer make a distinction between knowledge of vocabulary and knowledge of syntax – both types of language knowledge are involved in grammatical knowledge. However, Purpura's (1999) work gives evidence for an indeterminate distinction between vocabulary knowledge and syntactic knowledge. Schmitt (2000) further argues that knowledge of vocabulary and knowledge of syntax can be regard as "partners in synergy, with no discrete boundary" (p. 14). Nonetheless, Bachman and Palmer's (1996) classification of language knowledge indicates that an effort has been made to identify components underlying language knowledge. In addition, it provides as a useful reference point with which researchers can have a better understanding of what aspects of language knowledge their studies concentrate on. Therefore, in the current study, Bachman and Palmer's (1996) categorization was adopted. Following what several previous studies have done, (e.g., Bossers, 1991; Kobayashi, 2002; Purpura, 1997; 1998b; 1999), the current study restricted language knowledge to knowledge of vocabulary (termed "lexical knowledge" in the current study) and knowledge of syntax (termed "grammatical knowledge" in the current study).

2.8.2.2 Strategic competence

In Bachman and Palmer's (1996) model of language ability, strategic competence is developed based on that in Canale and Swain's (1980) model of communicative competence. Canale and Swain define *strategic competence* "as a set of compensatory strategies that could be used to overcome breakdowns or problems in communication" (Fulcher & Davidson, 2007:41). Clearly, strategic competence plays a compensatory role in the language use context. By contrast, within Bachman and Palmer's (1996) model of language ability, strategic competence is treated as higher-level cognitive processing in which cognitive behaviors are actively monitored, evaluated and managed to achieve a goal. It is composed of a goal-setting component, a planning component and an assessment component. In a language testing context, a goal-setting component concerns test-takers' capability for making a decision on what they are going to do, this decision which varies with test-takers' language knowledge, interests, demands of tasks, difficulty levels of tasks and so forth. A planning component is concerned with test-takers' capacity to take deliberate action to preview or overview designated tasks with an eye to establishing a general idea of how and when to do them appropriately with the use of available resources. Finally, an assessment component refers to test-takers' ability to make judgments against self-set criteria on what is necessary, how they accomplish given tasks, and how well they have completed them. This component often comes with a monitoring component thought of as purposeful action that test-takers take to supervise and check their cognitive processing or performance on given tasks. In so doing, confirmation and correction, suppose needed, are able to be made adequately so as to accomplish the given tasks successfully. It is evident that much of strategic competence is characterized as megacognitive capacities, "which underlie the way in which competence is related to performance" (Skehan, 1998: 166). Strategic competence in this model is not regarded as being "compensatory, only activated when other competences are lacking" (Skehan, 1998: 161). Rather, it is a crucial mechanism functioning all the time in actual communicative or language test situations.

Under some circumstances, for example, cognitively demanding settings (e.g., when considerable language-centered knowledge is required) or high-stakes situations (e.g., determining individuals' future study – a university entrance exam), the importance of strategic competence is highlighted (Phakiti, 2003). Within these contexts, in order to iron out difficulties experienced and optimize performance, test-takers consciously, intentionally and purposefully deploy so-called strategies which are manifestations of the activation and the operation of strategic competence. Notice that test-takers' language knowledge is related to the extent to which such activation and operation are profitable to completing given tasks. If test-takers lack a proper amount of language knowledge to rely on, the contribution of their applying and materializing strategic competence to task performance will be limited even though they invoke strategic competence.

2.8.2.3 Personal characteristics

According to Bachman and Palmer (1996), *personal characteristics* encompass elements such as age, gender, culture, attitudes, cognitive style and strategy use. Although not part of language ability, they are responsible for variances in task performance. Because of these components of personal characteristics, tests can not exactly measure what they are purported to gauge and the outcomes of tests can not completely account for the construct underlying tests. It seems impossible for test-designers to control this set of components thoroughly inasmuch as they stem from test-taker inherent attributes and vary across test-takers. Nevertheless, being aware of their existence is necessary. In this study, strategy use was centered on, given the motivation of the study (see Section 1.3 for details) and the feasibility of collecting related data.

In a test situation, test-takers with certain personal attributes (i.e., strategic test-takers) tend to accomplish tasks in a more strategic manner than others without. For example, as taking a multiple-choice reading comprehension test, they are likely to read entire test questions and alternatives first to get a general idea of test questions and make a prediction of the content of a reading passage. The understanding and the prediction serve as a frame of reference they can rely on when they process the passage and search for possible answers. They maybe outperform their counterparts with similar language knowledge on account of such strategy use. It follows that test-takers' performance on a given test is influenced by strategy employment to some extent and probably more than by the specific language ability that the test is originally intended to measure, as Bachman (1990) points out. It is worth noting that test-takers' strategy utilization may be inappropriate and counterproductive without the assistance of language knowledge and strategic competence through which strategy deployment is monitored and evaluated.

2.8.3 Test-taking strategies/processes in multiple-choice L2 reading tests

In an L2 reading test domain, substantial studies have shown that there are certain kinds of strategies existing, applied by test-takers during a test-taking course (e.g., Cohen, 1984; 1998a; 1998b; Cohen & Upton, 2006; 2007; Nevo, 1989; Rupp, Ferne & Choi, 2006). Among a diversity of reading test formats, a multiple-choice format has drawn a great deal of language testing researchers' attention by virtue of its unique nature – test questions, or stems and alternatives are provided. Aslanian (1985: 21) remarks that L2 test-takers are capable of arriving at correct answers to most of questions by the strength of clues available in test questions without an appropriate understanding "the meaning relationships, organization of the text, the reasoning pattern of the exposition, or what the text generally means, for that matter". However, this could be attributed to test construction, not to the test format itself. With cautious construction, multiple-choice reading tests still function to measure test-takers' reading ability to some extent, since test-takers need to make sense of reading passages or test questions to a certain degree in order to arrive at answers. The following study provides related evidence.

Dollerup, Glahn, and Hansen (1982: 96) in their preliminary study identified two types of the reading process in a multiple-choice reading test: *mainline reading* and *fragmented reading*. Mainline reading signified that test-takers skimmed the reading passage first and obtained the gist of it as a reference point during the reading process. On the other hand, fragmented reading denoted that test-takers consulted the words in the neighborhood of questions or the words with strong concepts, i.e. key words. These two pinpointed reading processes can be taken as an indication that test-takers attempted to comprehend the reading passage although in different ways. Further, they found that some test-takers made an educational guess that they arrived at answers based on their prior knowledge and clues emerging from alternatives. The finding that test-takers drew upon information on alternatives reveals a difference in test-takers' cognitive processing between in a multiple-choice reading test situation and in a regular reading context. For this, Cohen (1984) gives more evidence.

In Cohen's (1984) report, test-takers, in multiple-choice reading tests, tended to read the questions first instead of the passage, or read part of the passage and tried to look for a question related to it. These behaviors are more characteristics of reading in a test-taking situation than in a non-test-taking setting. In addition, Cohen quoted a study where test-takers were not given the passage on which multiple-choice items were based (Israel, 1982, cited in Cohen, 1984). Since there were four alternatives (one correct answer and three distractors), the chance of getting the correct answers would be 25 per cent, but the rate of success for test-takers at an advanced and an intermediate levels of language ability reached 49 and 41 per cent respectively. This suggests that they were drawing upon internal linguistic evidence rather than a wild guess procedure (McDonough, 1995). It follows that certain amount of language knowledge is essential to apply some strategies. A related finding is implied in Nevo's (1989) study.

Nevo (1989) found that the strategy *matching alternatives with the text* was employed with the high frequency, and more often in an L2 reading test than in an L1 one. Such matching entails test-takers' processing and understanding written texts as well as test items to a certain degree. Then, it can be argued that test-takers, when sitting L2 reading tests, need to rely on a certain amount of L2 language knowledge to be able to deploy this strategy appropriately. Substantively, she also found that test-takers' strategy use was related to the difficulty level of items – more demanding the items were, more non-contributory strategies (e.g., guessing or selecting the exception) were employed. More related findings are available in Anderson, Bachman, Perkins and Cohen's (1991) study.

Anderson *et al*. (1991) conducted a study to investigate the test-taking process that ESL test-takers underwent while taking a reading comprehension test and to relate the information to the content of reading comprehension test items as well as to their performance on those items. Anderson *et al.* found that in multiple-choice reading tests, the use of some strategies was significantly associated with either item difficulty or item types classified as main idea, inference and direct statement. This indicates that test-takers, when taking multiple-choice reading tests, appeared to vary the deployment of these strategies with test items at discrepant levels of difficulty or different item types. Another important finding was that test-takers monitored their reading comprehension during the test-taking process. The role of metacognitve awareness in test-taking processes is manifested and empirical evidence for its prominence in test-taking processes is showed in Anderson's (1991) and Farr, Pritchard and Smitten's (1990) studies.

Anderson (1991) reported a finding that the level of L2 proficiency contributed more to performance on a reading comprehension test than did the use of processing strategies. The finding is similar to that in L1-L2 reading research. Partially consistent with the implication in Cohen's (1984) report and Nevo's (1989) study, the author also found that strategy use was subject to part of language knowledge (i.e., vocabulary knowledge). This finding suggests a language threshold for deploying strategies appropriately. Anderson concluded that a marked discrepancy in strategy use between good performers and poor performers consisted in good performers being more aware of how to deploy strategies in an appropriate and flexibly manner. Anderson's conclusion underscores the importance of metacognitive awareness in testing-taking processes.

Analyzing the data collected through observations and introspective/retrospective interviews, Farr, Pritchard and Smitten (1990) found that compared with reading strategies employed while test-takers were primarily reading the passages, test-takers more frequently utilized question-answering strategies deployed while test-takers were primarily answering questions. The finding shows test-takers' tendency to use such strategies once test questions have been consulted. They also found that the most frequently used test-taking strategy was that of looking back in the passage to search for plausible answers. The finding agrees with that in Nevo's (1989) study. Farr *et al.* concluded that a multiple-choice reading comprehension test was a special reading task in which expert test-takers manipulated background knowledge, scanned, skimmed and reread the partial passage, reflected on the options, and postponed making a choice until they felt confident of a plausible answer. Then, it can be argued that the test-takers were prudent and reflective readers employing monitoring and meaning-oriented strategies. This illustrates that metacognitive awareness carries weight in a test-taking process.

With what has been discussed, we have learned that some strategies deployed in normal reading contexts are also employed in multiple-choice reading test-taking settings and metacognitive awareness is present not simply in normal reading process but in multiple-choice reading test-taking processes as well. Nonetheless, some studies discussed above still reveals variances in how L2 test-takers (readers) process texts in a multiple-choice reading test setting and a non-test reading context. The reason why language testing researchers strive to locate the discrepancies lies in an attempt to uncover what multiple-choice reading tests assess. This is concerned with the construct of reading comprehension of tests in a multiple-choice format. The following studies give an overview of this information.

From a perspective of an interaction between test-takers' meaning construction and test tasks, Gordon and Hanauer (1995) found that test tasks interacted with the test-takers' meaning development during reading tests. Given that multiple-choice reading tests provide test questions and options, more information is available and integrated into test-takers' meaning construction. It follows that the interaction between multiple-choice reading tests and test-takers' meaning development differs from that between regular reading tasks and readers' meaning construction to some extent.

A qualitative study conducted by Rupp, Ferne and Choi (2006) showed that L2 test-takers conceived responding to multiple-choice reading tests as a problem-solving task rather than a comprehension task particularly when meanings of distractors were very similar and plausible. This finding echoes Farr *et al.*'s (1990). Rupp *et al.* argued that response processes induced by multiple-choice reading tests varied from those by regular reading tasks. They also found that the construct of reading comprehension of a multiple-choice test involved different representations manifested through the characteristic of items, which broadened the definition of reading comprehension. Another critical finding by Rupp *et al.* was that some of test-takers' strategy deployment relied on their perceived characteristics of texts and test items. This can be taken as an indication that metacognitive awareness operated in the response process.

Similar to Rupp *et al.*'s (2006) finding, a recent qualitative study carried out by Cohen and Upton (2006; 2007) also indicated that L2 test-takers approached the new TOEFL reading section as a test-taking task – the priority was to arrive at correct answers rather than to learn anything from the passage read. Additionally, they found that some test-taking strategies were employed along with other test-taking or reading strategies (e.g., *consideration of options in context before a final decision is made* and *making a preliminary (but uncertain) selection of an option*). The finding supports the notion that strategies work in a combination manner, which has also been evidenced in L2 reading strategy research.

In summary, the research works discussed thus far cast light on how test-takers tackle multiple-choice reading comprehension tests. They take advantage of information emerging not only from texts but also from test questions and alternatives, and combine this with their own knowledge sources to reach a plausible answer. In a sense, the reading process in a multiple-choice reading comprehension test setting differs from that within a non-test reading context in certain ways. However, similar to strategic readers, strategic test-takers are aware of their test-taking processes in multiple-choice reading tests. They deploy strategies to process texts or arrive at answers, and monitor and evaluate their strategy employment. In this respect, the similarity is shared between a multiple-choice reading comprehension test setting and a non-test reading context. Additionally, they deploy their strategies in a way that one strategy or strategies is combined with another. This is also supported by findings yielded from other reading test situations. For example, in Nikolov's (2006) report, the analysis result of L2 test-takers' think-aloud protocols revealed that test-takers employed metacognitve, cognitive, and affective and social strategies in concert with one another. Finally, while strategy use may facilitate their performance, it appears that test-takers need to be equipped with a certain amount of L2 language knowledge and then they can apply some strategies appropriately to solve reading problems encountered or reach an answer in an L2 reading test.

Distinct from the previous studies discussed, the following quantitative studies look at test-taking processes from a metacognitive and cognitive strategy use perspective. Purpura (1997; 1999) adopted a structural equation modeling approach to examine the relationships between test-takers' strategy use and their performance on an L2 test (all subtests related to a reading test). He found that metacognitive strategy use exerted a direct and positive effect on cognitive strategy use. The finding lends empirical support to the concept that metacognitive strategies function as an executor for cognitive strategies. However, surprisingly metacognitve strategy use had no direct effect on L2 test performance. Nonetheless, it influenced L2 test performance indirectly via cognitive strategy use. This implies that cognitive strategies operate in tandem with metacognitive strategies. Another key finding was that the use of memory strategies had a detrimental effect on performance on a lexico-grammatical test. The finding suggests that strategy deployment does not always contribute to test performance. Purpura concluded that whether test-takers benefited from strategy use relied on how they took advantage of strategies and employed a strategy along with other strategies to tackle various tasks.

Purpura (1998b; 1999) further investigated the effect that strategy use exerted on high- and low-ability test-takers' L2 test performance. Several differences were located. For instance, unlike that in the high-ability group, metacognitive strategy use exerted a significant total effect on performance on all subtests in the low-ability group. Further, the high-ability test-takers depended on self-evaluating strategies to pay attention to formal features of the language; thereby, they could perform well on grammar, vocabulary, and cloze subtests. However, the self-evaluating strategies displayed no effect on the low-ability test-takers' performance. Finally, the high-ability test-takers, on the whole, reported utilizing strategies less frequently than the low-ability test-takers except for five strategies: monitoring, inferencing, self-evaluation, practicing naturalistically and linking with prior knowledge. Although this study offered substantive information on strategy use variations across groups with different L2 ability, participants' not referring to tasks when they filled in the questionnaire may lead to the collected data somewhat being unreliable.

To address this drawback, Phakiti (2003) conducted a study in which participants took a reading test and then completed a strategy use questionnaire. She found differences in the use of cognitive and metacognitive strategies across test-takers who were unsuccessful, moderately successful and highly successful readers. The finding indicates that test-takers' strategy use varies with their reading ability in the L2 reading test context. Importantly, she also found that compared with those who were moderately successful readers, test-takers who were highly successful readers were more metacognitively aware of their strategy use during the reading test. Phakiti argued that it was metacognitive awareness that allowed test-takers who were highly successful readers to outperform those who were moderately successful readers on the reading test. She concluded that strategic competence in Bachman and Palmer's (1996) model of language ability should not be simply operationalized by metacognitive strategies. While the findings here are valuable and informative, the role of L2 language knowledge in L2 reading test performance is not taken into account.

Predicated on Purpura's (1999) work, Phakiti (2008) applied a structural equation modeling approach to look at the relationship of test-takers' trait strategy use and state strategy use to their EFL reading test performance over time. In accord with Purpura's (ibid.) findings, Phakiti's study indicated that metacognitive strategy use impacted upon cognitive strategy use which yielded a direct effect on L2 reading test performance over time.

To summarize, the studies reviewed above make it clear that test-takers with high reading ability partially differ from those with low reading ability in metacognitive and cognitive strategy deployment. Both metacognitive and cognitive strategy employment display effects on reading test performance. However, echoing findings in L2 reading strategy research (e.g., Padron & Waxman, 1988; Sarig, 1987), strategy deployment does not guarantee to facilitate reading test performance. Metacognitive strategy deployment also yields effects on cognitive strategy deployment. In addition, compared with cognitive processing, metacognitive awareness is more inferential in reading test-taking processes because it dictates more test performance.

2.9 Conclusion

This chapter provides a series of literature reviews. Each of this research has contributed a great deal to our understanding of reading and test-taking processes. Reading and test-taking strategies are employed during reading or test-taking processes; nonetheless, the deployment of these strategies is not necessarily beneficial to reading performance or comprehension of what has been read. L2 proficiency/L2 language knowledge is related to, and exercises an influence on strategy use in the L2 reading context. Strategy deployment is not completely the same across language ability levels. L2 language knowledge, no doubt, plays a vital role in L2 reading test performance. However, some methodological limitations are available in previous studies.

Firstly, qualitative studies can demonstrate an effect of strategy use on reading test performance, based on the evidence that readers invoke strategies to deal with their incomprehensible parts, and the incomprehensible parts are solved by their employment of strategies. Nevertheless, little information is shown about the strength of the effect that readers' strategy use yields on their reading test performance.

Secondly, within some research works (e.g., Oxford, Cho, Leung, & Kim, 2004; Phakiti, 2003), questionnaires, rather than being generated from similar participants, are adapted from other studies, which leads to the fact that the validity and reliability of the collected data may be compromised. To explain, strategy use is subject to users and tasks (Grenfell & Harris, 1999; Nevo, 1989; Rupp *et al*., 2006). Given the limited similarities in participants and given tasks between their studies and other studies, participants in their studies are likely to employ strategies which are not listed on the questionnaires administered. Then, the data regarding the use of these strategies can not be gathered. It follows that the validity and reliability of the collected data perhaps are impinged upon.

Thirdly, some previous studies utilize questionnaires to elicit strategy use without allowing participants to make reference to a task (e.g., Mokhtari & Reichard, 2004; Purpura, 1997; 1998b; 1999; Sheorey & Edit, 2004; Sheorey & Mokhtari, 2001). One criticism for this is that the elicited strategy use is not overtly reliable, seeing that participants may overestimate or underestimate their strategy use in a task-absent situation. In order to minimize this drawback, the call for task-based strategy assessment arises (Cohen, 1998b; Hsiao & Oxford, 2002; Oxford *et al*., 2004) – participants' strategy use is collected immediately after they complete a given task.

Fourthly, within some studies (e.g., Oxford, *et al*., 2004; Padron & Waxman, 1988; Phakiti, 2003), the psychometric characteristics of questionnaires or tests are not examined. Then, little information is available on whether the construct validity of a strategy use questionnaire or a test is present.

Finally, among most previous relevant studies (e.g., Anderson, 1991; Barnett, 1988; Davis & Bistodeau, 1993; Oxford, *et al*., 2004; Padron & Waxman, 1988; Phakiti, 2003), data analysis methods are limited to analysis of variance, multivariate analysis of variance, regression analysis, or correlation analysis. These analyses are useful in demonstrating that strategy use is related to L2 proficiency or strategy employment varies to some extent across tasks or groups with different L2 proficiency. They are also conducive to revealing whether strategy use predicts task performance and to what extent. However, because of the inherent limitations, they fail to provide a whole picture of the relationship among English language knowledge, strategy use, and reading test performance with effect paths among these components in a single modeling framework.

All of these limitations are related to the methodology of my study which will be discussed in the next chapter.

CHAPTER THREE
METHODOLOGY

3.1 Introduction

In Chapter One, I posed the following research questions, given what has been discussed in Section 1.3.

1. What is the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance?
   1.1 Do students' English language knowledge and reading and test-taking strategy use contribute to their multiple-choice reading comprehension test performance? If yes, what are their relative contributions to multiple-choice reading comprehension test performance?
   1.2 Is there a language threshold for students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance?
   1.3 What is the relationship between students' English language knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test?
2. Is there a difference in the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance across English ability levels?
   2.1 Is there a difference in students' English language knowledge and reading and test-taking strategy use contributing to their multiple-choice reading comprehension test performance across English ability levels? Do the relative contributions of students' English language knowledge and reading and test-taking strategy use to their multiple-choice reading comprehension test performance differ across English ability levels?
   2.2 Is there a language threshold for students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance across English ability levels?
   2.3 Is there a difference in the relationship between students' English language

knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test across English ability levels?

In order to find answers to these research questions and overcome some methodological limitations of previous studies, I adopted a quantitative research approach. The research design was an ex post-facto correlational design utilizing survey methodology. Prior to data collection, a retrospective interview with participants and reading tasks similar to those in the formal study was conducted to develop the strategy item pool for a strategy use questionnaire (see Section 3.6.2.2). The verified measures – an English language knowledge test, a strategy use questionnaire and a multiple-choice reading comprehension test – functioned as instruments. 1064 EFL students from six senior high schools located in the south region of Taiwan participated in the study. Their strategy use was collected immediately after they completed the reading test. Structural equation modeling (SEM) was applied to examine the relationship among students' English language knowledge, strategy use and their reading test performance.

In the following sections, I first present a brief conceptualization of survey research. Then, I discuss my research design and the nature of measurement. Next, I describe instrumentation, participants, data collection procedures, statistical techniques for data analysis. Finally, I conclude this chapter with an outline of the pilot study.

3.2 Survey research

Surveys, traditionally, acquire data at a particular point of time with a purpose of (a) depicting the conditions that have existed and identifying standards against which existing conditions can be compared; (b) determining the existing relationships between specific events (Cohen, Manion, & Morrison, 2001). The former is a *descriptive survey*, whereas the latter is an *explanatory survey*, which is what Oppenheim (1992) refers to as "the analytic, relational type of survey" (p. 12). More specifically, a descriptive survey concerns the frequency of the occurrence of an event investigated and the number of people who have a certain opinion about, or take a certain attitude toward an event of interest. However, it neither explains anything nor reveals causal relationships. Different from a descriptive survey, an explanatory or analytic survey is concerned with providing explanations of an event investigated and looking for the relationship of particular variables.

In addition, according to the length of data collection, survey research is categorized into *cross-sectional* and *longitudinal* surveys (Babbie, 2004; Wiersma & Jurs, 2005). In cross-sectional surveys, data is collected from a sample only once at a particular time. It is a one-shot study; thus, the change within the sample can not be captured. Different from cross-sectional surveys, data in longitudinal surveys has been gathered from the same sample over a period of time. The development or change of research issues is the concern of this type of survey research.

As Morrison (1993) puts it, there are a number of positive features of surveys. To begin with, since surveys allow data to be collected on a one-shot basis, they are efficient and economical. Also, the data produced is number-oriented and capable of being processed in a statistical manner. Accordingly, surveys offer both descriptive and inferential information. Additionally, in a survey study, vital factors are manipulated and standardized information is collected by means of uniform instruments for all participants. Finally, seeing that a survey study generally hinges upon large data gathered from a wide population, the result is more representative and more likely to be generalized to other context[2].

Despite the positive features mentioned above, survey research, like other research approaches, has its limitations. First of all, unique or unexpected events or instances concerning research issues perhaps are unable to be identified. The survey research is also difficult to provide comprehensive explanations for issues explored. Further, the context in which research questions are involved fails to be depicted deeply and exhaustively. Finally, there is a slight chance of researchers' portraying in detail the change or development of research issues over time.

Within survey research, questionnaires, standardized tests, attitude scales, and structured or semi-structured interviews are often employed to collect data. Analysis of variance, multivariate analysis of variance, regression analysis or correlation analysis are performed to understand the group differences in variables of interest, the amount of variance that a variable accounts for another variable, the most important or less important determinants, or the relationship among variables. It is worth noting that although a cause-and-effect relationship can be explored in the explanatory/analytic survey research, the identified relationship among variables in survey studies does not denote the same cause-and-effect relationship as that in an experimental study. As Mertens (1998) observes, in the survey research, no treatment variables are

---

[2] Although survey research, traditionally, is conducted on a large-scale basis, a small-scale survey research is still permissible – the generalizability of the findings that are produced from this type of survey research is limited.

experimentally manipulated; as a consequence, an identified causality relation cannot be viewed as definite proof of a cause-and-effect relationship.

In the present study, an explanatory/analytic and cross-sectional survey was carried out in six senior high schools located in the south region of Taiwan to investigate the relationship among senior high school students' English language knowledge, strategy deployment, and their reading comprehension test performance. With the application of a multivariate analytic procedure, structural equation modeling (SEM), an attempt was made to identify cause-and-effect relationships approximating those in an experimental study.

3.3 Research design

This study adopted an ex post-facto correlational research design. Students' English language knowledge and strategy use were assessed after they completed a reading test. Then, the relationship among students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance was investigated by using a multivariate analytic method. Further, whether the aforementioned relationship varied across groups with different English ability was also examined. English language knowledge was limited to *lexical knowledge* and *grammatical knowledge*. *Lexical knowledge* referred to students' vocabulary breadth, whereas *grammatical knowledge* related to knowledge students had of syntactic rules, prepositions and word usage. *Reading and test-taking strategies* referred to the conscious and/or subconscious mental and behavioral activities that affect student performance on multiple-choice reading comprehension tests—either directly or indirectly. *Reading and test-taking strategy use* related to the deployment of these strategies. *Multiple-choice reading comprehension test performance* referred to how well students performed a multiple-choice reading comprehension test intended to measure reading for main ideas, facts, or details of reading passages. It also refers to how well they draw inferences.

A hypothesized model of the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance was specified by employing a sophisticated statistical methodology – *structural equation modeling* (SEM). SEM can test the postulated relationships among English language knowledge, strategy use, and reading comprehension test performance in a single model framework. Effects that English language knowledge and strategy use have on reading test performance can be calculated by a set of mathematical equations and shown in an accepted model. Students' English language knowledge was gauged by an English language knowledge test, whereas their reading and test-taking strategy use was assessed by a reading and test-taking strategy use questionnaire. Their multiple-choice reading comprehension test performance was measured by a multiple-choice reading comprehension test. The framework of the research design is illustrated in Figure 3.1.

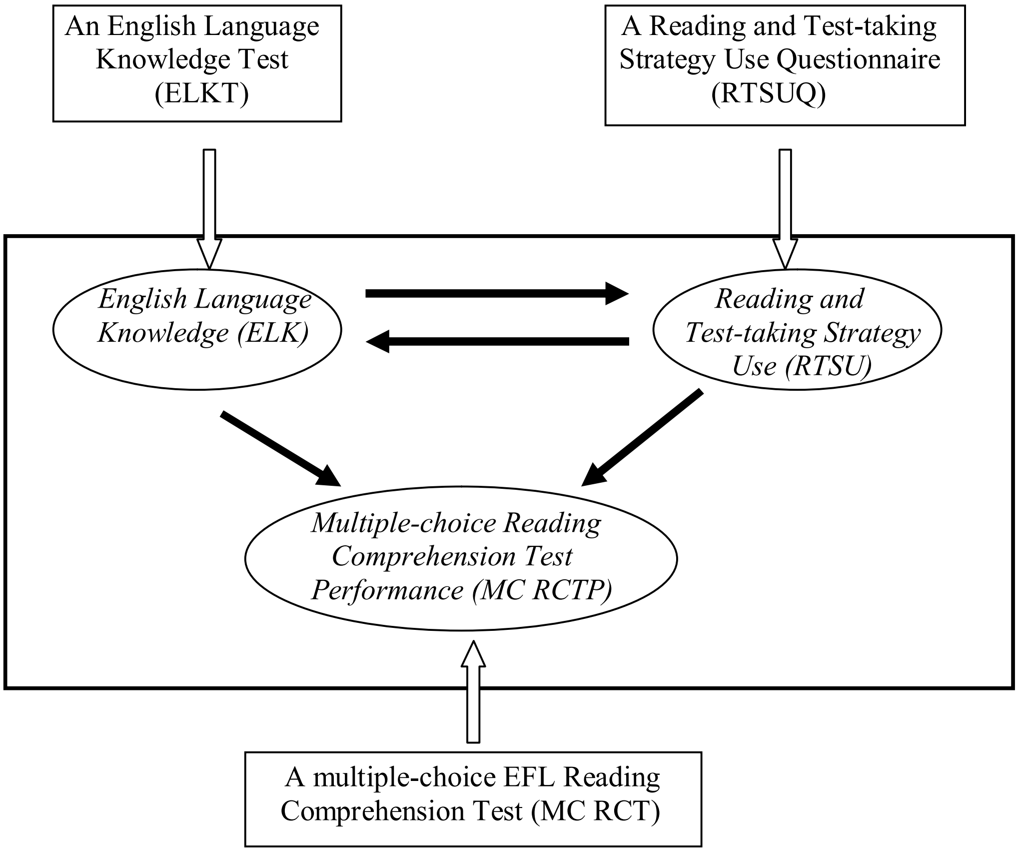


Figure 3.1 The framework of the research design

In the central part of Figure 3.1, there is a large rectangle. This characterizes what the current study aims to look at: the relationship among English language knowledge, strategy use, and reading test performance. Inside the large rectangle are three horizontal ovals representing factors examined in the current study: English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance. A thick, bold, single-headed arrow "➡" symbolizes an effect that one factor yields on another. The three small rectangles respectively refer to an English language knowledge test, a reading and test-taking strategy use questionnaire, and a multiple-choice reading comprehension test. These measures function to assess the factors investigated in the present study and a single-headed arrow “ ⇨ ” represents this relationship.

3.4 Measurement

As stated in the previous section, an English language knowledge test, a strategy use questionnaire and a reading test were utilized to respectively assess students’ English language knowledge, strategy use and reading test performance (for details see 3.6.2). Therefore, it is worth briefly discussing what measurement is. *Measurement* refers to a process during which a set of standardized systems is adopted to describe observed objects or unobserved concepts (Nunnally & Bernstein, 1994). In survey research, measurement plays a pivotal role since it allows data to be efficiently collected in quantity and contributes to organizing the sizeable amount of gathered data. Furthermore, with the aid of measurement, some light can be cast on individuals’ underlying perceptions, attributes, or behaviors of interest. Based on the results of measurement, statistical analysis can be run and inferences can be drawn to answer research questions posed. However, scores obtained from measurement fail to deeply manifest participants’ unique individual responses to different facets of research issues – a general picture is what measurement principally provides. In addition, scores obtained from such a measure as an attitude scale show the strength of an attribute, but they can not reflect the whole story underlying the strength. These limitations are related to the properties of measurement.

Measurement features properties of indirectness, incompleteness and imprecision. It is conceivable that scores obtained from measurement just indirectly represent attributes or behaviors of interest. By means of the scores, attributes or behaviors can be interpreted and understood. However, gained scores are indicators of partial, rather than entire, attributes or behaviors under investigation. It is a myth to claim that what has been acquired by measurement reveals the exhaustive picture of attributes or behaviors investigated. The complete insight into individuals’ attributes or behaviors is, in fact, difficult to get. All we can do is assume that what is obtained is representative of what is of interest. Given that scores acquired from measurement symbolize only a portion of attributes or behaviors, then measurement deviates from accuracy to a certain extent. There is also very little likelihood of constructing a measure which contains items all equally representative of attributes or ability (Bachman, 1990). In addition, during the measurement process, other unobserved factors are involved and impact on the measurement result more or less. Accordingly, measurement is not as precise as it is expected to be.

Given what is discussed above, in the current study, caution would be taken when the measures were administered, results were interpreted, and inferences were made. After all, what was obtained from an English language knowledge test, a strategy use questionnaire and a reading test represented part of students' English language knowledge, strategy use and their reading test performance.

3.5 Validity and reliability in terms of measures

During the measurement process, validity and reliability are important issues in that they have a crucial impact upon the quality of collected data. Given that the present study utilized measuring instruments to collect related data, it is worth addressing validity and reliability in terms of measures. For measures, *validity* focuses on the degree that they are able to gauge what they really aim to measure, while *reliability* concerns consistency – a measure behaves "in a fashion which is consistent with itself" (Oppenheim, 1992: 159). A fair measure is supposed to feature a certain level of validity and reliability. Reliability is highly correlated with validity. A measure with an inadequate degree of reliability fails to possess an appropriate degree of validity. A reliable measure, on the other hand, is not necessarily a valid one. There are a number of types of validity and reliability regarding measures. In the current study, *face validity*, *content validity*, *construct validity* and *internal reliability* are focused on and addressed.

Face validity signifies that measures seem to assess what they are intended to gauge. The more they appear to assess the targeted attribute, the more face validity they are equipped with. Face validity can be evaluated by expert judgments – the experts with related knowledge are invited to appraise whether contents and test items or response items of measures are adequate. The degree of the agreement of their judgments functions as an indicator of the appropriateness of contents and test items or response items of measures.

Content validity is concerned with the degree of appropriateness and representation of test items or response items. Content validity of measures can be achieved through ascertaining that test items or response items of the measures are appropriately related to the assessed ability or attribute and the sampling of test items or response items is representative enough that the target ability or attribute can be gauged. Expert judgments mentioned above can also be adopted to ensure content validity of a measure.

Measures are supposed to assess underlying constructs that represent what is intended to gauge. *Construct validity* of measures hinges upon the extent to which test items or response items function as suitable indicators of underlying constructs or concepts. It is the most difficult type of validity that evidence can be proffered for (Seliger & Shohamy, 2000). When scores gained from a measure highly correspond to underlying constructs that a measure is constructed to assess, construct validity of the measure exists.

A reliable measure denotes that all test items or response items in the measure assess the targeted attribute accurately and consistently. The more reliable a measure is, the more closely a true estimate of the attribute assessed by the measure can be reached (Mertens, 1998). Two common approaches are available to estimate reliability: *repeated measures* (*test-retest, equivalent forms*) and *calculation of internal consistency*. The latter is widely reported in language studies on account of its unique strength. Through this technique, reliability of a single form of measure administered only once can be estimated, distinct from repeated measures, which require two administrations or two forms of measures. Cronbach's alpha (α), one of the most frequently reported reliabilities, is this type of reliability. The higher Cronbach alpha a measure obtains, the more reliably it functions.

In the current study, an expert judgment technique was utilized to assess face validity and content validity of the measuring instruments administered. In addition, they were submitted to exploratory factor analysis (EFA) in order to extract underlying components and to demonstrate the presence of construct validity. The reliability of the measures is determined by internal reliability with the computation of Cronbach's alpha coefficients. In the subsequent section, what is related to the data collection will be discussed.

3.6 Data collection

Data was collected from June to August, 2007. Originally, eight senior high schools located in the south region of Taiwan were selected and contacted. However, English teachers in two schools expressed that they were unable to provide enough time for the data collection; consequently, data collection was conducted in six senior high schools during English classes. In the following subsections, participants will be focused on at first, then measuring instruments, and finally data collection procedures.

3.6.1 Participants

Based on convenience sampling, 1064 EFL students from six senior high schools located in the south region of Taiwan served as participants in the current study. All of them were third-graders, aged 17-18. As the data was gathered, they were preparing for the Senior High Academic Ability Examination, held in the coming February. As a result, they were expected to be experienced in test taking, which could be conducive to collecting related data. Their first language was Chinese. At the time of the study, they had been learning English as a foreign language at least for five years. After invalid tests and questionnaires were dropped, the final sample ended up with 834. Table 3.1 provides an overview of their background information.

Table 3.1 Background information of participants in the current study

Grouping	N	Grouping	N	Total N
Gender				
Male	630	Female	204	834
School				
PD	410		0	410
PDG	0		42	42
FS	8		26	34
FH	92		75	167
KO	105		0	105
CS	15		61	76
Year of English learning				
Between five to ten years	588		184	772
Above ten years	42		20	62
Going to cram school				
Yes	577		188	765
No	53		16	69
	Mean		Mean	Total M
Self-rating English ability	11.405		11.623	11.458

Note. PD, PDG, FS, FH, KO and CS were pseudo-names for the schools where data was collected.

As shown in Table 3.1, the number of the male participants amounted to 630 (76%), whereas that of the female participants corresponded to 204 (24%). The number of the male participants was three times larger than that the female participants.

Appropriately half of the participants came from one school (PD), with participants 410 (49%). Most of the participants had learned English for five to ten years (n = 772, 93%). The majority of the participants went to cram school for English learning (n = 765, 92%). The result comes as no surprise since the coaching phenomenon is common in Taiwan and most Taiwanese senior high school students usually go to cram school to better their English ability after class. As for self-rating English ability, the mean for the overall participants (11.458) exceeded the half of the total score (20) simply by a little bit. This suggests that participants appeared to take a conservative attitude towards self-rating their English ability. Furthermore, there was no difference in the mean for self-rating English ability across gender groups.

3.6.2 Instrumentation

In the present study, three types of instruments were administered: (a) an English language knowledge test; (b) a reading and test-taking strategy use questionnaire; and (c) a multiple-choice reading comprehension test. A pilot study was conducted in September and October, 2006 to see how well the tests and the questionnaire worked. Some test items and strategy items were deleted based on the results of item analysis. This will be explained in detail in Section 3.8. In the following, an English language knowledge test will be centered on at first, then a reading and test-taking strategy use questionnaire, and finally a multiple-choice reading comprehension test.

3.6.2.1 An English language knowledge test

According to Bachman and Palmer (1996), *language knowledge* consists of organizational knowledge and pragmatic knowledge. Then, *organizational knowledge* is made up of grammatical knowledge and textual knowledge; *pragmatic knowledge* is comprised of functional knowledge and sociolinguistic knowledge. Although language knowledge may not be demarcated as definitely as they describe, their categorization proffers a general picture of what components language knowledge comprises. Theoretically, all these aspects of knowledge should not merely be measured, but be assessed by authentic tasks and four language skills should be included with an eye to obtaining a complete picture of participants' language knowledge. However, it appears not plausible to do so in the present study, given a large number of the participants (N = 1064) and the limited resources available. Additionally, Bachman and Palmer (1996) note that often simply one or a few aspects of language knowledge are focused on and measured by language tests developed. Consequently, given the feasibility of the current study, it is necessary to select areas of language knowledge to be centered on, despite the fact that doing so will limit the constructs of language knowledge.

Lexical and grammatical (syntactic) knowledge, as part of language knowledge, plays a crucial role in L2 reading performance. In Hoover and Gough's (1990) simple view of reading, Rumelhart's (1977) interactive model of reading, or Stanovich's interactive-compensatory reading model (1980), reading comprehension is likely to be impeded if readers lack sufficient lexical knowledge to process printed words efficiently. In Auerbach and Paxton's (1997) and Yorio's (1971) studies, L2 learners claim that deficiency of lexical knowledge is the chief source resulting in their L2 reading difficulty. Previous L2 reading research works also suggest that lexical knowledge is essential and influential in L2 reading to some extent (e.g., Droop & Verhoeven, 2003; Laufer & Sim, 1985; Nassaji, 2003b; Shiotsu & Weir, 2007; van Gelderen, Schoonen, de Glopper, Hulstijn, Simis, Snellings & Steveson, 2004). On the other hand, L2 reading studies have been supporting the notion that L2 grammatical/syntactic knowledge gives certain weight in L2 reading comprehension (e.g., Barnett, 1986; Cohen *et al*., 1979; Devine, 1988; Laufer & Sim, 1985; Nassaji, 2003a; 2003b; Paribakht, 2004; Shiotsu & Weir, 2007). Further, a number of researchers contend that L2 learners' lexical and grammatical/ syntactic knowledge exerts an effect on their L2 reading performance (e.g., Barnett, 1986; Grabe, 1991; Koda, 2005; Schulz, 1983; Shiotsu & Weir, 2007). Accordingly, in the present study, language knowledge was limited to lexical knowledge and grammatical knowledge. Lexical knowledge referred to students' vocabulary breadth, whereas grammatical knowledge related to knowledge students had of syntactic rules, prepositions and word usage.

Within previous research concerning L2 or FL reading (e.g., Bernhardt & Kamil, 1995; Bossers, 1991; Kobayashi, 2002; Lee & Schallert, 1997; Pichette *et al*., 2003; Taillefer, 1996; Usó-Juan, 2006), vocabulary and grammatical subtests were administered to assess participants' language knowledge or language proficiency. With what has been carried out in these studies being followed, within the current study, English language knowledge, with an acknowledgement of its limitation, was operationalized by an English language knowledge test, which consisted of grammatical and vocabulary subtests. The English language knowledge test originally comprised fifty-five grammatical test items and sixty vocabulary test items. After validation (see the fourth and the fifth paragraphs on p. 64 for how the test was validated), twenty-nine items for the grammatical subtest and twenty-seven items for the vocabulary subtest were retained, with a Cronbach's alpha (α) of .915 for the validated English language knowledge test (see Appendix 1 for the English language knowledge test used in the current study).

As far as the grammatical subtest was concerned, I designed test items by myself. The reasons for this were the following. To begin with, I could not have access to standardized grammar tests. In addition, these tests might not appropriately fit participants' language proficiency. They could be so challenging for participants as senior high school students that participants perhaps make a guess to a certain great extent while sitting the grammatical test. Then, the validity and reliability of the test will be diminished. Given these reasons, I constructed the grammatical subtest on my own by making reference to textbooks published in 2005 by Sanmin, Far East, as well as Longtung publishers and used in senior high schools in Taiwan. Grammatical test items were designed to measure the following grammatical concepts: nouns[3], pronouns, tense, mood, participles, adjectives, infinitives, gerunds, adjective clauses, noun clauses, adverb clauses, inversion clauses (verbs), conjunctions and prepositions. According to the curriculum syllabi issued by the Ministry of Education in Taiwan (Ministry of Education in Taiwan, 2006), these grammatical notions should be covered by the three-year formal English instruction at the senior high school level. In the grammatical subtest, each test item provides a sentence or two sentences with a part supplanted by a blank. Four options are given, among which only one meet the syntactic constraints imposed by the structure of the rest of the sentence or context offered. An attempt was made to minimize the involvement of the processing of sentence semantics to ensure that the test was as valid as impossible, although this was not that easy. An example for grammatical test item is " _____ her work, Susan took a rest under a tree. (A) Finished (B) Had finished (C) Having finished (D) She finished".

With respect to the vocabulary subtest, there were two test sections included: the definition matching section and the sentence completion section. I constructed test items in the definition matching section. Eighteen measured words were selected according to the Vocabulary List for Senior High School Students issued by the Ministry of Education in Taiwan (College Entrance Examination Center in Taiwan, 2006c). The vocabulary items in the list are classified into six levels: Level One is the lowest, while Level Six is the highest. For instance, "admit" is set at Level Three, while "borrow" is set at Level

---

[3] Test items which functioned to assess the grammatical concept regarding nouns were deleted after item analysis.

Two. Within the definition matching section, there were three subsections. All measured words were verbs in the first subsection; nouns in the second subsection; adjectives in the third subsection. The words utilized to describe the definitions of measured words were mostly at the lower level than the measured words were at. For example, the measured word "affection" is at Level Five and the words used to describe its definition (a feeling of liking or love and caring) are at Level One or Two, lower than the level of the measured word. Only few words were at the same level as the measured words were at. The overall section consisted of measured words ranging from Level One to Level Six. Within this test section, participants were required to choose a definition for a measured word from a word definition bank provided.

Within the second section, test items were drawn from the vocabulary test section of an English test of the Senior High Academic Ability Examination as well as the counterpart in the Senior High Appointed Subject Examination in Taiwan from 2002 to 2006 (College Entrance Examination Center in Taiwan, 2006a; 2006b). With a view to making the test more valid, several test items were revised so that the level of words used in the items was largely lower than that of measured words. On no account was the level of words used in the test items higher than that of the measured words. In this section, each test item provided a sentence or two sentences with a part replaced by a blank. Four options were offered, among which only one satisfied the semantic constraints imposed by a sentence or sentences. In an effort to minimize the involvement of grammatical components in the vocabulary test, the sentence structure of a test item was simplified as much as possible. An example for test items in this section is "If you want to borrow magazines, tapes, or CDs, you can visit the library. They are all _____ there. (A) marvelous (B) available (C) sufficient (D) impressive".

In order to assure the quality of the English language knowledge test, the following principles were observed.

Firstly, each test item had only one answer. Among four options, only one option was correct when it was placed in a blank given by a test item.

Secondly, a test item assessed nothing more than one feature at a time.

Thirdly, with a view to making distractors plausible and attractive, each option was grammatically correct when placed in a blank of a test item, except options in a grammatical subtest.

Fourthly, the length of all options was kept approximately equal for fear that a correct option was too obvious or distractors malfunctioned.

Fifthly, the level of test items was set appropriately. Most of the test items were at a lower level than measured words. For example, the measured word “temporary” was at Level Three according to the Vocabulary List for Senior High School Students and the words used in the test item “Mr. Smith’s work in Taiwan is just ____ . He will go back to the U.S. next month” were at level One lower than the level of the measured word.

Sixthly, test items, on the one hand, were maintained as clear as possible to express sufficient information; on the other hand, as brief as possible in order not to bore participants – the length of most test items was within fifteen words.

Seventhly, less demanding test items were placed at the initial part of the test, followed by more challenging ones. In the vocabulary subtest, a test item with a measured word at Level One was placed at the initial part of the definition matching section. A test item with a measured word at Level Three was placed at the initial part of the sentence completion section.

Eighthly, the test was evaluated by an assistant professor who taught in the department of applied foreign languages in a university in Taiwan and a lecturer with a PhD degree in Education who taught in the department of modern languages in a university in the UK to make sure content validity and face validity of the test were appropriate. They both agreed that content validity and face validity were present in the test.

Finally, the English language knowledge test was piloted (see Section 3.8 for details) and item analysis was carried out to delete some test items. Both the item discrimination index method and the point-biserial correlation were performed for item analysis. Exploratory factor analysis was also conducted to give evidence for the fact that the construct validity of the English language knowledge test is present to a certain degree. Additionally, the internal reliability was calculated to ensure that this test functioned as a reliable measure. The result indicated that this English language knowledge test functioned reliably ($\alpha = .915$).

3.6.2.2 A reading and test-taking strategy use questionnaire

Techniques typically utilized to gather data to understand the nature of strategies are comprised of observations, interviews, verbal reports, diaries and journals, and self-report questionnaires. Among these methods, a self-report questionnaire is the most frequently adopted and efficient technique to understand learner strategy use (Chamot, 2005; Cohen, 1998b; Oxford, 1996; White, Schramm, & Chamot, 2007). In the field of

L2 reading or language testing, a host of studies are available in which questionnaires are administered to look at reading/test-taking processes that L2 readers/test-takers go through or reading/test-taking strategies that they deploy (e.g., Barnette, 1988; Carrell, 1989; Hong-Nam & Leavell, 2006; Macaro & Erler, 2008; Oxford *et al.*, 2004; Padron & Waxman, 1988; Phakiti, 2003; 2008; Purpura, 1997; 1998b; 1999; Mokhtari & Reichard, 2004; Sheorey & Mokhtari, 2001). In the current study, a six-point Likert-type scale questionnaire was utilized to examine Taiwanese senior high school students' reading and test-taking strategy use when they sat a multiple-choice reading comprehension test (see Appendix 2 for the strategy use questionnaire administered in the current study).

In order to contribute to a better understanding of the questionnaire used in this study, it is worth addressing briefly what a Likert-type scale is. A typical Likert scale is comprised of a set of response items (i.e., statements) constructed to measure attributes or behaviors of interest (Dörnyei, 2003). Response items are allocated with an array of continuous numbers which represent the degrees that respondents agree or disagree with what response items state. The scores of all response items are usually summed or averaged. The final score indicates the degree of opinions or attitudes under investigation (e.g., very much → not at all, or strongly agree → strongly disagree).

A Likert-type scale questionnaire features several advantages in terms of the current study. Firstly, "the [collected] data are more uniform and standard" (Seliger & Shohamy, 2000: 172), which is conducive to the subsequent quantitative analysis. Further, since the strategy data obtained from Likert scales are continuous scores with variances, the structural equation modeling analysis can be performed to answer research questions posed in the present study. Additionally, according to Robson (1993), the structure of Likert scales appeals to respondents. This contributes to participants in the current study being willing to respond to strategy items. It follows that in their response processes a certain degree of consideration is involved, which exerts a positive effect on the validity and the reliability of elicited data. Moreover, compared with other techniques, Likert-type scale questionnaires appear to allow participants to self-report their strategy use more easily. Participants do not need to verbalize their intricate mental states and perform a given task simultaneously, as in the case of think-aloud procedures.

On the other hand, there are several drawbacks to a Likert-type scale questionnaire. First of all, it may be demanding for participants to decide a number that exactly represents the extent to which they agree or disagree with strategy items. They may underestimate or overestimate the degree of their agreement or disagreement. In addition, the extreme point on the scale may be avoided because of human beings' common wishes – appearing like others in many aspects (Cohen, Manion, & Morrison, 2001). Hence, a neutral number on the scale may be preferred. Finally, participants perhaps misinterpret strategy items on the questionnaire, which diminishes the validity and reliability of collected data.

In order to minimize the limitations stated above, several steps were taken. For example, participants filled out a Likert-type scale questionnaire immediately after completing a reading test that functioned as an elicitor. In so doing, the possibility that participants felt difficulty in determining a number to covey their opinions on strategy items might be reduced. Further, during the data collection process, participants were encouraged to express their actual viewpoints on strategy items by informing that their responses had nothing to do with their academic records and would be treated confidentially. In addition, the questionnaire was piloted and some items were deleted or revised prior to the formal administration of the questionnaire to ensure that the meanings of strategy items were appropriate. In the data collection course, participants were also given a Chinese version of the questionnaire and allowed to ask questions if they did not understand the meanings of strategy items. In the next section, the focus shifts onto how the strategy use questionnaire was developed.

The reading and test-taking strategy use questionnaire used in the current study was constructed, predicated on Pressley and Afflerbach's (1995) *model of constructively responsive reading*, and Rogers and Bateson's (1991; 1994) *model of test-taking behavior of skillful test-takers*. Pressley and Afflerbach's (1995) model of constructively responsive reading has been adopted by Mokhtari and Reichard (2002) to develop their metacognitive awareness reading strategies inventory. The concept of constructively responsive reading is built on detailed analyses of a sizeable number of protocols and seems to be compatible with such a recognized reading theory as Rosenblatt's (1978) *reader response theory* in which the transaction between readers and texts is accented. In addition, the notion of constructively responsive reading subsumes a cluster of key components of a *bottom-up oriented text-processing approach* submitted by van Dijk and Kintsch (1983), a *top-down oriented text-processing approach* – schema theory – advocated by Anderson and Pearson (1984), as well as *comprehension monitoring processes* in which evaluation is often involved – metacognitive theory – proposed by Baker and Brown (1984). Apart from what is stated above, the conception of constructively responsive reading also encompasses *inference-drawing processes* put forward by Graesser & Kreuz (1993). In summary, the model of constructively responsive reading features the following:

(a) readers seek overall meaning of text, actively searching, reflecting on, and responding to text in pursuit of main ideas;
(b) readers respond to text with predictions and hypotheses that reflect their prior knowledge;
(c) readers are passionate in their responses to text;
(d) readers' prior knowledge predicts their comprehension processing and responses to text (Pressley & Afflerbach, 1995: 99-102).

This model consists of a number of processes in which readers are engaged in to make sense of the text they read and provides more insights into reading processes based on substantial empirical evidence from L1 reading. It follows that specifically speaking this model is referred to as L1 reading process model. When this model is applied to an L2 reading context, this model may not be so appropriate since L2 reading is more complicated. However, this model is still preferred and adopted here because it is grounded on abundant empirical evidence from self-reports. In a sense, it is reliable to a certain degree.

As for Rogers and Bateson's (1991; 1994) model, this model is predicated on several researchers' work (e.g., Brown, 1980; 1987; Flavell, 1979; Schuell, 1986) and a preliminary model put forward by Smith (1980, cited in Rogers & Bateson, 1991) about how expert test-takers arrive at possible answers in a multiple-choice test. Additionally, Rogers and Bateson's model is supported by their own study (1991; 1994) in which senior high school students served as participants, which is similar to the current study. Their proposed model indicates that the cognitions of expert test-takers are composed of the following elements: (a) a cognitive monitor which functions to determine what capabilities are going to be involved in the question-answering process and orchestrate these abilities to reach a plausible answer to a question being addressed; (b) capabilities as well as knowledge pertinent to what is being measured; (c) test-wiseness knowledge[4]; (d) the response to items, including a procedure for choosing a possible answer and a record of the selection process. The model is presented in the following figure.

---

[4] Test-wisenenss is attributed to the characteristics of test methods and test-takers. It is defined as ability that test-takers have to enhance test performance by taking advantage of test formats, characteristics of tests or test contexts.

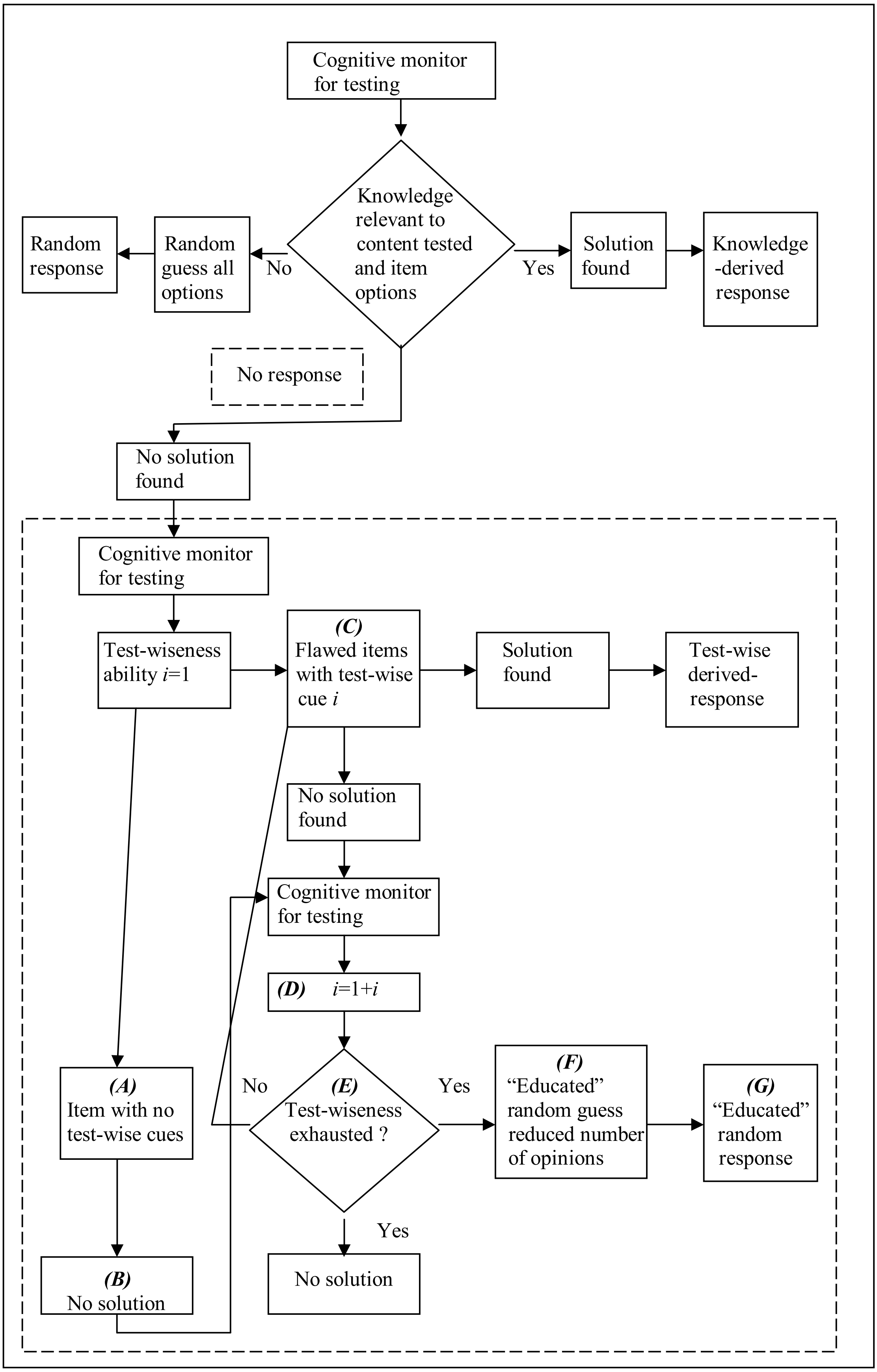


Figure 3.2 The Model of Expert Test-takers' Test-taking Behavior (Rogers & Bateson, 1991: 332)

As illustrated in Figure 3.2, the upper part of the large rectangle concerns how test-takers obtain an answer based on content knowledge assessed, the knowledge which is symbolized by a diamond-shaped rectangle. Two rectangles at the right of the diamond-shaped rectangle represent that an answer is obtained. By contrast, two rectangles at the left of the diamond-shaped rectangle characterize that an answer is not found and a random guess is made. Within this area, test-wiseness ability is not tapped into.

Alternatively, the inner broken line rectangle is the area where test-wiseness capability is applied to work out a possible answer (see Figure 3.2). At the top left hand corner of the broken line rectangle, there are two small rectangles which signify the cognitive monitor operating in the answer-searching course and test-wiseness ability (*i*) that test-takers dawn upon – " i =1" means the current test-wiseness capacity activated and applied. If an item is not equipped with test-wise cues, then related test-wiseness ability will not be invoked to answer a question, as shown in the two rectangles labeled *(A)* and *(B)*. (For the ease of reference, I labeled some rectangles.) Suppose an item happens to have a test-wiseness cue (*i*), as shown in the rectangle tagged *(C)*, then a match is made between test-wiseness ability (*i*) applied and an item with a test-wiseness cue (*i*) and an answer is arrived at. If not, the test-wiseness capability employed in the first time will serve as a frame of reference and be combined with the test-wiseness ability activated and applied next time, which is shown as "i =1+ i " in the rectangle labeled *(D)*. Test-wiseness may be exhausted, which is represented by a diamond-shaped rectangle tagged *(E)* and then an educated guess is made, as indicated by two rectangles labeled *(F)* and *(G)*, to get to an answer.

Within this model, the question-answering process is defined as the following paths. First of all, test-takers read and understand multiple-choice test items and options provided and then pick out a possible answer from given options, by means of knowledge about what is measured. Less skilled test-takers will guess randomly and choose an answer without any reasoning or just leave the item unanswered suppose they fail to arrive at a possible answer. Distinct from less skilled test-takers, expert test-takers, with the cognitive monitor, employ partial knowledge about the content gauged, information emerging from test items and options, as well as the set of test-wiseness principles to work in a cyclical way via "the elements of the set for a test-wiseness element-item cue match" (Rogers & Bateson, 1991: 333) until they reach plausible answers. The cycle of matching comes to an end when a match is made and such test-wise response is documented. There is greater likelihood that expert test-takers make an educated guess if no match is made, which may be because no items exist that test-wiseness principles can be applied to or test-takers run out of their test-wiseness strategies.

This model illustrates that the cognitive monitor plays a pivotal role in the overall question-answering process. Moreover, from this model it is obvious that both what is intended to measure and expert test-takers' attributes come into play in their performance on multiple-choice questions. Test-takers actively not merely draw upon their own cognitive resources, but information available from a given task to optimize their chances of arriving at correct answers. Although not profiling the entire picture of test-taking processes that test-takers engage in as they sit a test, this model at least depicts how expert test-takers reach a possible answer.

In order to develop a strategy-item pool for a reading and test-taking strategy use questionnaire, I conducted a retrospective interview in late March and early April, 2006. Based on convenience sampling, twelve second-graders were selected as participants from a senior high school in Taiwan, which was also one of the schools where formal data collection was carried out. All of the participants were 17-year-old male students and their first language was Chinese. In addition, they had been learning English as a foreign language at least for five years.

Participants first took a multiple-choice EFL reading comprehension test which consisted of three reading passages and twelve test items drawn from the reading comprehension test section of an English test of the Senior High Academic Ability Examination (SHAAE) used in 2003, 2004 and 2006. Participants were informed to heed their mental or behavioral activities that occurred as they sat the reading test. Thirty minutes was allocated for the whole test. Every time when completing test items of a reading passage, participants would say "stop." The test time was suspended and the interview was conducted immediately in Chinese. Participants were asked to move back to the beginning of the reading passage and test items. They read the passage as well as test items again, and recalled and reported their mental or behavioral activities that took place when they processed the reading passage and answered the test items. After the interview was finished, the test time was restarted and participants continued to tackle their next reading passages and test items. This process was repeated until participants completed the last reading passage or the test time ran out. The entire interview process was tape-recorded and completed within fifty minutes.

I adopted the reading/test-taking strategy taxonomy shown in Anderson *et al.*'s (1991), Nevo's (1989), Pritchard's (1990) and Yamashita's (2002) work as a starting point to code and classify strategies identified. Additional strategies reported by participants in the retrospective interview, but not listed in the above-mentioned work, were categorized and added to the list of strategies. Eighty-six strategies were located with six categories: monitoring, supporting, global processing, local processing, compromising, and test-taking. A PhD candidate specializing in reading strategies and I independently double-checked the 25% of translated transcription, identified strategies and the final categorization. A final agreement about different opinions was met through discussions. The inter-reliability of coding was .794, the result which was acceptable. With the result of the retrospective interview and the work of several researchers (e.g., Anderson *et al*., 1991; Cohen, 1998b; Nevo, 1989; Phakiti, 2003; Yamashita, 2002), strategy items for a strategy use questionnaire were then developed.

The questionnaire originally consisted of eight-five strategy items. After validation, seventy-two items remained (see the first and the third paragraphs on p.72 for how the questionnaire was validated). The original questionnaire was written in English and then translated into Chinese. In order to avoid participants' English proficiency impinging upon their filling in the questionnaire, participants received the questionnaire in Chinese. The questionnaire was roughly divided into three sections (see Appendix 2). The first section stated the purpose of the questionnaire and included the directions about how to respond to strategy items. The second section contained strategy items. The possible responses that participants would circle would be 0, 1, 2, 3, 4 or 5. The third section involved basic information about participants, such as their class, number, self-rating English ability, how long they have learned English, or whether they have been to cram school to learn English. In so doing, a general picture of participants' background information would be gained and an understanding of what types of participants from whom data was collected would be provided.

Allan's (1995) study indicates that a checklist (similar to a questionnaire) exerts an instrument effect on collecting self-reported strategies and "that it biased the responses, introducing random error" (p. 151). In order to enhance the validity and reliability of this data collection technique, the following steps were adopted.

Firstly, an introductory statement was written and set at the beginning of the questionnaire to explain the purpose of the study and encourage participants to fill in questionnaires deliberately and honestly.

Secondly, the introduction section was immediately followed by strategy items. Such arrangement was in order to facilitate participants' recalling their reading/test-taking strategy deployment in a multiple-choice EFL reading comprehension test they just took. Moreover, personal questions were placed at the end of the questionnaire. Apart from the aforementioned reason, this arrangement also lessened the sensitive nature of personal questions, which lent itself to participants' answering these questions honestly.

Thirdly, a PhD candidate, specializing in reading strategies, and a lecturer with a PhD degree in Education, teaching in the department of modern languages in a university in the UK, were invited to evaluate whether the layout and the expression of response items were adequate. Additionally, they appraised whether face validity and content validity were present – they all agreed that face validity and content validity were present in the strategy use questionnaire.

Fourthly, strategy items were presented to participants in Chinese: participants' first language. In addition, the Chinese version of the questionnaire was checked by two teachers who taught Chinese at the senior high school level to ensure the appropriateness of the wording. Also, both the English version and the Chinese version of the questionnaire were examined by a teacher who used to teach Chinese at the high school level and was doing her PhD degree in language education in the United States.

Finally, the questionnaire was piloted with the similar population in September and October, 2006 for item analysis (see Section 3.8 for details). Both the item-total correlation and the extreme group method were carried out for item analysis. Also, volunteers from amongst these participants were asked to reflect on the extent to which they understood the strategy items on the questionnaire. Some items were deleted. Further, with a view to determining the underlying psychometric characteristics of the questionnaire, it was submitted to exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The internal reliability was also calculated to ensure that this strategy use questionnaire functioned reliably; the result suggested that this strategy use questionnaire was a reliable scale ($\alpha = .953$).

3.6.2.3 A multiple-choice reading comprehension test

What two tests that aim to assess the equivalent ability gauge may vary with what test formats they are in. Reading tests are no exception. There is high likelihood that reading tests in different test formats measure diverse facets of reading constructs (Kobayashi, 2002; Urquhart & Weir, 1998). A number of test formats such as cloze tests, gap-filling tests, multiple-choice tests, constructed response tests, free recall tests, and summary tests have been developed and administered to access reading ability. Among these, a multiple-choice format probably is the most prevalently used method to assess reading ability (Alderson, 2000; Koda, 2005). In previous L2 reading studies, a plethora of researchers utilized multiple-choice questions to assess L2 reading comprehension (e.g., Block, 1986; Brantmeier, 2005; Bügel & Buunk, 1996; Carrell, 1989; 1991; Droop & Verhoeven, 2003; Lee & Musumeci, 1988; Lee & Schallert, 1997; Nassaji, 2003b; Rupp *et al*., 2006; van Gelderen *et al*., 2004). Despite offering retrieval cues that contribute to recalling what has been processed (Bransford, 1979), which may diminish the validity of a test and possibly not encouraging circumspect global reading, this format is believed to make it possible to check all reading levels (the semantic and syntactic facets of the passage), the discourse level (cohesion and coherence connections amongst diverse parts of the passage), as well as the pragmatic level (an author's point of view) (Harrison, 1983). In the present study, a reading comprehension test in a multiple-choice format was operationalized to assess Taiwanese senior high school students' reading comprehension test performance, given that it was this test format that the current study focused on.

Defining *reading comprehension* is a thorny task (Alderson & Urquhart, 1984; Carrell, 1991) in that the comprehension process can not be observed directly. It simply can be accessed indirectly by means of tests from which the mechanism of the comprehension process is inferred and interpreted (Wolf, 1993). However, it is a challenge to measure it in a methodologically well-developed and truly informative manner (Johnston, 1983; Swaffar, 1988; Taillefer, 1996). Inferences from the test result of a test task represent only a portion of reading comprehension. Theoretically, disparate test tasks should be adopted to assess reading comprehension in order to provide comprehensive insights into reading comprehension. However, in practice, it seems impossible given a substantial amount of time and labor involved in the measuring process and limited resources available. Accordingly, it is necessary to define reading comprehension for the current study.

The objectives of English language instruction pertaining to reading at the senior high school level in Taiwan subsume the following reading skills: scanning or skimming passages for specific or general information, looking for main ideas, drawing inferences from reading passages and guessing unknown words from context (Ministry of Education in Taiwan, 2006). In the present study, reading comprehension connoted reading for main ideas, facts, or details of the text and for drawing inferences from the text, with an acknowledgement of its limitation. Given the diversity of the topic of and the length of the passage, six reading passages and twenty-three test items except one[5] were drawn from the reading comprehension test section of an English test of the Senior High Academic Ability Examination (SHAAE) from 2002 to 2006.

With a view to assuring the quality of the reading comprehension test**,** the following principles were conformed to.

First of all, there was only one answer to each test question. Test questions with ambiguous options were excluded. Participants chose one possible answer from four options. All distractors were plausible.

Secondly, answers to test questions were passage dependent. In no way did participants arrive at possible answers without referring to reading passages.

Thirdly, the length of options was approximately equal. The case that the length of an answer or distractors was too prominent was avoided.

Fourthly, the language levels of stems or questions and of options were set at the lower than or the same as that of reading passages for fear that students' language knowledge might prevent them from making sense of test questions and options. Most words utilized in stems, questions and options were at Level Three according to the Vocabulary List for Senior High School Students issued by the Ministry of Education in Taiwan.

Fifthly, the order of reading passages placed on the exam papers was determined by the following criteria – a less challenging reading passage with the smallest number of total words was placed at the initial portion of the whole reading comprehension test, followed by more demanding ones.

Sixthly, prior to the administration of the reading test, the test was evaluated by an assistant professor teaching in the department of applied foreign languages in a university in Taiwan and a PhD candidate specializing in reading strategies to make sure content validity and face validity of the test were appropriate – both of them agreed that content validity and face validity were present in this multiple-choice reading comprehension test.

Finally, the reading test was piloted (see Section 3.8 for details) and item analysis was carried out to drop some test items. Both the item discrimination index method and the point-biserial correlation were performed for item analysis. Six reading passages and seventeen test items remained for the current study (see Appendix 3 for the reading comprehension test administered in the current study). Exploratory factor analysis was

[5] I constructed Item 17 in Passage F. The item was verified by an assistant professor who taught in the department of applied foreign languages in a university in Taiwan.

also conducted to provide evidence that the construct validity of the reading test was present to some extent. Additionally, the internal reliability was calculated to ensure that this reading test functioned as a reliable one. The result showed that this reading test was a reliable measure (α = .755).

3.6.3 Data collection procedures

Data collection was conducted from June to August, 2007. The data was collected in the classroom during English class sessions. Six senior high schools located in the south regions of Taiwan were chosen. Schools' and participants' consent for this study had been obtained in advance. Participants took the reading test and then filled out the strategy use questionnaire first. Three to seven days later, they sat the English language knowledge test. The directions were given in Chinese (participants' first language). Participants were encouraged to do their best, leaving no question unanswered and to sit the test as they did in the real test setting. How to fill out the answer sheet and the strategy use questionnaire was also explained. Further, it was emphasized that the result of the tests and the questionnaire would not be reported to teachers or school administrations and had nothing to do with their academic records. In addition, participants were cautioned not to disclose the contents of the tests and the questionnaire to others.

The reading test was issued to participants first. They were informed when test time was running out. Forty-five minutes were given for the reading test, more than enough time for nearly all of the participants, since it was not expected that participants rushed to complete the test and arrived at answers mostly on the basis of wild guesses. Prior to sitting the reading test, students were reminded to pay attention to how they approached the test, made sense of passages and arrived at plausible answers during their test-taking. Upon the completion of the reading test, participants received a strategy use questionnaire and moved on filling out the questionnaire. They were given twenty-five minutes to respond to strategy items on the questionnaire. With such an amount of time, participants were expected to recall their strategy use and respond to each strategy item carefully and honestly. As for the English language knowledge test, participants took the grammatical subtest first, and then the vocabulary subtest. They were given twenty minutes for the grammatical subtest and twenty minutes for the vocabulary subtest respectively. The directions given in the reading test mostly were also applied to this test.

The present study examined Taiwanese students' strategy employment after they completed a reading test. Such a data-collection procedure features several advantages.

First of all, students' reading and test-taking process is not disturbed and interrupted heavily when they are required to report their strategy use after sitting a test. Their overall reading comprehension test performance can be better understood because intrusion into the test-taking process is minimized. It also can be assumed that students' strategy deployment directly affects their performance in a natural manner.

Moreover, during the real reading process, students seldom verbalize their mental or behavioral activities; in the test-taking course, they are not allowed to utter a word. Thus, that strategy deployment is collected after the test allows the reading and the test-taking processes in the current study to be more compatible with those that occur in the genuine reading test context.

Finally, this study concentrated on investigating strategy use for the overall reading test. What is concerned with is the extent to which students agree or disagree with strategy items rather than the number of times they deploy a particular strategy for tackling the test. While filling in a strategy use questionnaire, students are assumed first to retrieve their strategy deployment from their working memory. Then, they make a judgment about their strategy use and convey the extent to which they agree or disagree with strategy items on a Likert-type scale. Such ecological factors regarding the overall test as the characteristics of reading passages, test questions and item difficulty are mirrored by this method (Phakiti, 2003).

Despite such advantages as mentioned above, I would bear in mind that gathered data pertinent to students' strategy use was limited to a reading comprehension test in a multiple-choice format.

3.7 Data analysis

The current study used the SPSS 15.0 (Statistical Package for the Social Sciences) statistical software package and the software AMOS 7.0 (Analysis of Moment Structures) for data analysis. Analyses performed in the present study were composed of descriptive statistics, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), single group structural equation modeling (SEM), a t-test, and multi-group structural equation modeling. A significance level of 0.05 ($p < 0.05$) was set. Nonsignificant results were reported by exact probability levels and indicated by "$p > 0.05$" and significant results were marked by "$p < 0.05$". Prior to explaining the major statistical analytic procedures, *exploratory factor analysis* (EFA), *confirmatory factor analysis* (CFA) and *structural equation modeling* (SEM) are briefly described.

3.7.1 Exploratory factor analysis (EFA)

As Jöreskog and Sörbom (1989, cited in Purpura, 1999) remark, *exploratory factor analysis* (EFA) is "a technique often used to detect and assess latent sources of variation and covariation in observed measurements" (p. 96). EFA functions as a useful analysis procedure in summarizing data with a small set of factors, uncovering the characteristics of collected data, or exploring interrelationships among an array of variables.

In the current study, EFA was applied to examine the construct validity of the reading comprehension test and the English language knowledge test, and to extract the components (constructs) of the questionnaire data. In addition, it was utilized as an initial step to identify the latent variables underlying the measuring instruments used for the subsequent construction of measurement models.

3.7.2 Confirmatory factor analysis

Distinct from EFA, *confirmatory factor analysis* (CFA) is an array of more sophisticated techniques used to confirm or disconfirm investigated hypotheses or theories regarding the structure underlying a set of variables. In the SEM analysis, CFA sets out with a postulated measurement model and then the model is accepted or rejected based on the model fit statistics and meaningful interpretations.

In the present study, CFA was utilized to examine the result produced from EFA for the strategy use questionnaire. It was also conducted to inspect the relationship between observed variables and latent variables for the measurement models of English language knowledge and of reading and test-taking strategy use, with the use of the SEM procedures.

3.7.3 Structural equation modeling (SEM)

Structural Equation Modeling (SEM) refers to a method which consists of several statistical analyses: confirmatory factor analysis, multiple regression analysis, analysis of covariance, and path analysis. As remarked by Bentler (1995), "linear structural equation modeling is a useful methodology for statistically specifying, estimating, and testing hypothesized relationship among a set of substantively meaningful variables" (p. ix). SEM, specifically speaking, "is a multivariate analytic procedure for representing and testing (a) inter-relationships between observed variables and constructs, and (b) inter-relationships among constructs" (Purpura, 1997: 300), predicated on theoretical underpinnings or previous empirical studies.

Within SEM analysis, *latent variables* (*constructs*) refer to attributes unable to be observed directly in the real world, whereas *observed variables* (*measured variables*) function as indicators of these latent variables. Each latent variable requires at least two indicators and each indicator is assumed to have an element of a measurement error. An SEM model which concerns the relationships between observed variables and latent variables is termed a measurement model. On the other hand, a model concerned with the relationships amongst latent variables is labeled as a structural model. A model comprising two or more measurement models and a structural model is called a full latent variable model.

According to Jöreskog (1993), SEM models are yielded respectively under the following three conditions: (a) *strictly confirmatory*; (b) *model comparison*; (c) *model generating*. In a *strictly confirmatory* condition, researchers construct a sole theoretical model and test this model with a set of collected data to determine whether the model is accepted or rejected. Within a *model comparison* condition, researchers, according to theory or empirical studies, construct several alternative models and test these models with gathered data to decide which model is the best. Finally, in a *model generating* condition, researchers construct a tentative model grounded on theoretical underpinnings or previous research, and then test this model with empirical data. If the model does not fit the data satisfactorily, the model is modified and respecified. The process is repeated until a generated model describes the data well.

In the present study, SEM was utilized in all three conditions. A strictly confirmatory procedure was used to test the relationship between observed variables and latent variables in the measurement model of English language knowledge. A model generating procedure was applied to test the relationship between observed variables and latent variables in the measurement model of reading and test-taking strategy use and construct the full latent variable model for the entire group and for groups with different English ability. Finally, a model comparison procedure was utilized to validate the strategy use questionnaire and in the simultaneous group analysis[6] to justify the appropriateness of the accepted model.

Generally, in the SEM procedures, a set of relationships between observed variables (measured variables) and latent variables (constructs) or among latent variables is hypothesized and specified in a model with the use of a cluster of mathematical

[6] In the simultaneous group analysis, two or more models are analyzed simultaneously to test whether parameters on the paths shared by these models are equivalent across these models.

equations. A hypothesized model is tested by evaluating the goodness-of-fit between the model and collected data. A set of model fit indices is adopted to appraise a hypothesized model. If the model fit statistics of the hypothesized model satisfy the requirements of these indices, which indicates that the model describes the data well, then the model is accepted. If not, the model is rejected or modified. A poor model is re-specified and retested until a final model with desirable goodness-of-fit and meaningful interpretations is yielded.

Similar to other statistical analytic procedures, SEM has several limitations. Firstly, the research findings yielded from SEM are based on a single hypothesized model that fits the collected data. However, there are still maybe a number of alternative models which may fit the data better (Dörnyei, 2007). As a result, the findings are tentative. Secondly, during the model producing process, despite the fact that the model modified describes the gathered data satisfactorily, over-reliance on the modification indices to modify the model may result in the model being meaningless and un-interpretable. Finally, although SEM can identify causal effects, these effects still should be interpreted cautiously. Causal effects identified by SEM do not equate to those pinpointed in an experimental study where variables are under careful control.

However, SEM features the following advantages that contribute to the current study investigating the relationship among students' English language knowledge, strategy use and their reading test performance. Firstly, SEM can analyze and present the relationship between observed variables and latent variables or the relationship amongst latent variables within a single modeling framework. Given this advantage, a clear picture of the relations amongst students' English language knowledge, strategy use and their reading test performance can be given. Secondly, an effect of a variable on another variable can be calculated by a set of mathematical equations. With this ability, effects that English language knowledge and strategy use have on reading test performance can be shown in an accepted model. Thirdly, SEM manifests more accurately what measures assess by providing measurement errors in the hypothesized model. Then, more understanding can be gained that an English language knowledge test, a strategy use questionnaire and a reading test, in fact, do not completely assess students' English language knowledge, strategy use and reading test performance. Fourthly, as mentioned above, SEM combines several analyses, such as confirmatory factor analysis or multiple regression analysis, which makes it possible to perform these analyses at a time within one hypothesized model. Finally, SEM allows the postulated relationship to be simultaneously examined across groups. Stated another way, whether the parameters of interest are invariant across groups can be inspected and tested in a way that these parameters are constrained to be equivalent across groups and estimated simultaneously. With such an advantage, the present study can supply more accurate findings regarding differences in effects of English language knowledge and strategy use on reading test performance across groups with discrepant English ability. Given these advantages, compared with other statistical analytic procedures, SEM is more powerful to analyze the relationship among students' English language knowledge, strategy deployment, and their reading test performance.

3.7.3.1 SEM assumptions

SEM has several assumptions which should be met with an eye to obtaining trustworthy results.

Firstly, the assumption of a reasonable sample size should be satisfied or at least approximated in order to ensure the accuracy and stability of estimates and the representativeness of the outcome. SEM researchers typically recommend that the larger the sample size, the better. However, a large sample size affects some of model fit indices. A balance should be made. When reviewing the literature, Shumacker and Lomax (1996) found that the sample size ranged from 200 to 500 in most SEM research. As recommended by Stevens (1996), and Bentler and Chou (1987), minimum 15 cases for per observed variable should be an acceptable criterion. A sample size less than 150 may compromise the external validity and not ensure stable estimates (Kunnan, 1998). In the present study, the sample size was appropriate in both the entire group analysis and the separate group analysis. Based on Stevens's (ibid.), and Bentler and Chou's (ibid.) criterion, the minimum sample size for the current study with 18 observed variables was 270 (15×18). For the entire group analysis, the sample size ended up with 834. For the separate group analysis, the sample size of the high English ability group was 312, while that of the low English ability was 522. All the sample sizes exceeded the minimum sample size required.

Secondly, most estimation procedures adopted for SEM assume that data is not simply univariately but multivariately normally distributed as well. An examination of the skewness and the kurtosis of each observed variable can see whether the univariate normality assumption is satisfied. As for multivariate normality, the skewness and the kurtosis for all observed variables can illustrate whether the multivariate normality assumption is met. The univariate normality assumption is satisfied when the multivariate normality assumption is met (Hung, 2002). In the current study, the multivariate normality was inspected by an assessment of normality provided by the AMOS software. If data was distributed non-normally greatly, several cases (possible outliers) were deleted to ensure that the multivariate kurtosis[7] value was within the accepted limits[8] and this assumption was not violated too much (see Section 4.3.1 for details).

Finally, SEM features the linearity assumption. A linear relationship means that the relationship of two observed variables forms a straight line. In addition, it denotes that a new variable, after the linear combination of a set of variables, correlates to other variables linearly. However, when a host of observed variables are involved in a study, it is difficult to see whether the linearity assumption is satisfied. To explain, among a set of variables, a pair of them may be found to be related to each other linearly by observing the scatter plot of the two variables. However, this process becomes complex when more variables are involved. Consequently, the conceptual meaning matters rather than the practical meaning of the linearity assumption (Chiu, 2006). In the current study, it was this conceptual meaning regarding the linearity assumption was taken.

3.7.3.2 Evaluation of overall model fit

Selecting appropriate indices in evaluation of a hypothesized model is one of the most demanding tasks pertinent to SEM analysis. Due to the absence of a single unanimously recognized criterion (Heubeck & Neil, 2000) and the recommendation of combining several indices being made (Bollen, 1989; Sasaki, 1993; Schumacker & Lomax, 1996; Tseng, Dörnyei, & Schmitt, 2006; Vandergrift, Goh, Marechal, & Tafaghodtari, 2006), a number of commonly accepted model fit indices were adopted to evaluate a postulated model in the current study.

The chi-square statistic (χ^2), although presented in the final report of the current study, is not used to evaluate the goodness-of-fit of a model in that it is strongly influenced by the sample size (Byrne, 2001; Floyd & Widaman, 1995; Marsh, Balla, & McDonald, 1988; Wu & Tu, 2005). The chief indices adopted to appraise a hypothesized model are as follows: the *goodness-of-fit index* (*GFI*), the *adjusted goodness-of-fit index* (*AGFI*), the *comparative fit index* (*CFI*), the *Tucker-Lewis index* (*TLI*) as well as the *root*

---

[7] Kurtosis refers to the peakedness of a distribution of the data. In the SEM analysis, the value of the multivariate kurtosis functions as an indication of whether the data is distributed multivariately normally.
[8] Kline (1998) suggests that when the absolute value of kurtosis is over 10, then a distribution of the data is regarded as a non-normal distribution. In the current study, this criterion was adopted.

mean square error of approximation (*RMSEA*). These indices have been used in previous strategy-related or L2 studies in which SEM is applied (e.g., In'nami, 2006; Phakiti, 2008; Purpura, 1997; 1998b; 1999; Sasaki, 1993; Schoonen, Hustijn, & Bossers, 1998; Shiotsu & Weir, 2007; Tseng *et al.*, 2006; Vandergrift *et al.*, 2006). In addition, the finding in Marsh, Balla and Hau's study (1996) lends support to the adoption of the CFI and the TLI in the evaluation of the model fit. The previous studies also indicate that the RMSEA outperforms other indices as an index of appraising the model fit (e.g., Browne & Arminger, 1995; Marsh & Balla, 1994). As claimed by Rayhov (2001), this index has served as a well-informed indicator of the overall evaluation of the model fit. The following provides a general idea of these model fit indices.

Firstly, the *goodness-of-fit index* (*GFI*), sensitive to the sample size, corresponds to the R square in multiple regression analysis. This index represents the extent to which variances and covariances of a hypothesized model could explain variances and covariances of the collected data. The range of the index is from 0 to 1, with the value above .900 being desirable.

Secondly, the *adjusted goodness-of-fit index* (AGFI), influenced by the sample size, corresponds to the adjusted R square in multiple regression analysis. This index is used to compare the goodness-of-fit of different models within the same data or the goodness-of-fit of the identical model for discrepant groups. The index ranges from 0 to 1, with the value greater than .900 being acceptable.

Thirdly, the *comparative fit index* (CFI) is determined by comparing a hypothesized model with the independence model in which observed variables do not correlate with each other. It proffers a full measure of covariance in the data. In addition, as pointed out by Bentler (1990), this index, ranging from 0 to 1, depends little on sample size. The acceptable value is greater than .950 (Hu & Bentler, 1999).

Fourthly, the *Tucker-Lewis index* (TLI), also known as the non-normed fit index (NNFI), indicates the difference between a hypothesized model and the independence model in which observed variables do not correlate with each other. The index is likely above 1 and the acceptable value is greater than .950 (Hu & Bentler, 1999).

Finally, the *root mean square error of approximation* (RMSEA), affected by the number of estimated parameters in the model (Byrne, 2001), reveals the extent to which a hypothesized model varies from the saturated model, the model which fits the data perfectly. A value of this index less than .060 indicates acceptable model fit (Hu & Bentler, 1999).

3.7.3.3 AMOS (Analysis of Moment Structures) notation and terms

In the current study, the software AMOS 7.0 (Analysis of Moment Structures) was used to perform SEM. In AMOS graphics, circles represent latent variables (constructs) or residuals, whereas squares symbolize observed variables (measured variables). A residual associated with observed variables is named as *E* (*measurement error*, also called *uniqueness*). A residual related to latent variables is labeled as *D* (*disturbance*). Bidirectional arrows "⟷" signify *correlations* and *covariances* between variables without a defined casual direction. Single-headed arrows " ⟶ ", in contrast, represent *factor loadings* in a measurement model or *standardized regression coefficients* (*effects*) in a structural model, showing a causal effect that one variable exerts on another. In addition, direct effects refer to those that one variable yields directly on another, while indirect effects represent those that one variable, by means of other variable(s), displays indirectly on another. Total effects encompass direct effects and indirect effects.

3.7.3.4 Statistical identification of models

In the SEM analysis, a hypothesized model should be identified at first and then parameter estimation can be performed. Identification, broadly speaking, is concerned with "whether or not there is a unique set of parameters consistent with the data" (Byrne, 2001: 35). The model identification centered on the extent to which a unique array of values can be inferred for the unknown parameters, based on a given covariance matrix of analyzed variables. Specifically speaking, there is at least one unique solution for parameter estimation in an SEM model. If a model has only one possible solution for parameter estimation, then it is a *just-identified* model. In this type of the model, the number of variances and covariances of observed variables equals the number of parameters to be estimated. In this situation, no degree of freedom is present since no difference exists between the number of variances and covariances of observed variables and the number of parameters to be estimated. Supposing in a model there are an infinite number of possible solutions for parameter estimation, then the model is called an *under-identified* model. Within this model, the number of variances and covariances of observed variables is less than the number of parameters to be estimated. The input data feeds the model with insufficient information. In this case, a model can not be tested. Finally, provided a model has more than one possible solution for parameter estimation, the model is regarded as an *overidentified* model. In this model, the number of variances and covariances exceeds the number of parameters to be estimated. It is this model with positive degrees of freedom that is preferred in the SEM analysis.

In the current study, AMOS presented an identification problem by no parameter estimate being shown in a model after the model was estimated, signifying that the model was underidentified. As a result, supposing a model was estimated and parameter estimates could be revealed, it was assumed that it was a just-identified or overidentified model.

3.7.4 Data analysis procedures

In the current study, data analysis procedures, as shown in Figure 3.3, consisted of descriptive statistics, exploratory factor analysis, confirmatory factor analysis, single-group structural equation modeling, a t-test and multi-group structural equation modeling.

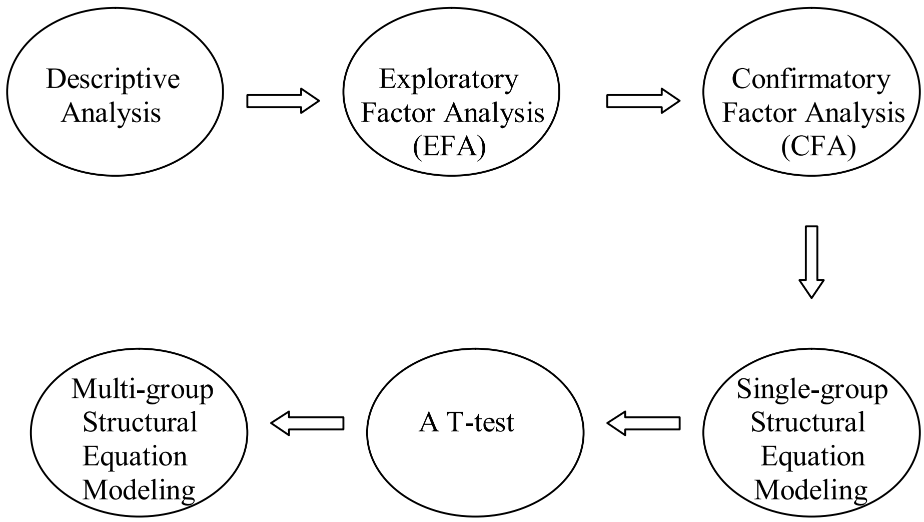


Figure 3.3 A flow chart of statistical procedures used in the current study

First of all, descriptive statistics was performed for the English language knowledge test, the strategy use questionnaire and the reading comprehension test. Test items or strategy items of these measures and the entire measures were described in the light of the average score (means), variability (standard deviations) and distribution of scores (skewness and kurtosis). Reliability of these measures was examined by computing the internal consistency.

Secondly, a series of exploratory factor analyses (EFAs) was performed to extract constructs (components) underlying these measures. Then, based on the results of EFAs, the separate measurement models were proposed of English language knowledge, of reading and test-taking strategy use, and of multiple-choice reading comprehension test performance. Sections 4.2.1, 4.2.2, and 4.2.3 provide these models.

Thirdly, confirmatory factor analysis (CFA) was conducted on the measurement model of English language knowledge and that of reading and test-taking strategy use to examine the relationship between latent variables and observed variables in these measurement models, with the use of SEM. The reading comprehension test was not submitted to CFA because it was simply followed by two indicators. The limited number of observed variables in this measurement model resulted in CFA not being run on this test, with the application of SEM. *Maximum likelihood estimation* procedures, a type of widely used estimation procedures to estimate parameters in the SEM analysis, were adopted due to its statistical robustness. The finding in other studies shows that only when the absolute value of the multivariate kurtosis is larger than 25 will parameter estimates obtained by maximum likelihood estimation procedures be influenced (Chiu, 2006; Hung, 2002; Muthén & Muthén, 1998). Prior to performing SEM, the *multivariate normality assumption* was examined, the assumption which should not be violated as the maximum-likelihood estimation procedures are adopted.

Fourthly, SEM was also applied to analyze the relationship among students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance.

Fifthly, participants were divided into two groups, based on the results of their self-rating English ability: the High English Ability (HEA) group and the Low English ability (LEA) group. Next, an independent samples t-test was conducted to test whether there was a difference in English language knowledge, reading and test-taking strategy use and multiple-choice reading comprehension test performance respectively between these two groups. This analysis functioned as an initial step to locate cross-group discrepancies and justified the appropriateness of the subsequent multi-group analysis.

Finally, two analyses were carried out for the multi-group analysis: *the separate group analysis* and *the simultaneous group analysis*. For the separate group analysis, a full latent variable model for the HEA group and for the LEA group was respectively generated by means of SEM. Then, a comparison was made between these two groups to pinpoint commonalities and differences in the component structures of English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance, and in the structure of the relationships among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance.

With regard to the simultaneous group analysis, the HEA group model and the LEA group model were estimated simultaneously with equality constraints being imposed on the parameters for the paths shared by the two groups. Cross-group equality constraints were released one by one, and a produced model was evaluated, based on model fit indices and critical ratios for difference between parameters. This process was repeated until all cross-group equality constraints on parameters of interest were examined.

3.8 The pilot study

The pilot study was conducted in September and October, 2006. The data was collected in the classroom during English class sessions. Based on convenience sampling, the participants for the pilot study were chosen from eight different classes of a senior high school in the south region of Taiwan. Four classes at the second-grade level were selected, whereas four classes at the third-grade level. After invalid questionnaires and tests were excluded, the final sample ended up with 283. All of them were male students, aged from 16 to 18. They shared similar linguistic, culture, and socioeconomic backgrounds. These students sat a reading comprehension test and then filled in a strategy use questionnaire. A week later, they took an English language knowledge test.

With the collected data, at first, I performed an item analysis to drop some unsatisfactory test items or strategy items. As for the reading comprehension test and the English language knowledge test, with Wu and Tu's (2005) suggestions being followed, the item discrimination index method and the point-biserial correlation were used for item analysis. With respect to the item discrimination index method, an item was accepted if it could discriminate well between the total test scores of the upper 33 percent and the lower 33 percent of the participants. The .250 cut-off was adopted in this analysis. As for the point-biserial correlation, an item was acceptable if the correlation coefficient between it and the scale were .300 or above. Finally, the reading comprehension test consisted of six reading passages and seventeen test items, with the appropriate internal reliability ($\alpha = .755$). The English language knowledge test was composed of twenty-nine test items for the grammatical subtest and twenty-seven test items for the vocabulary subtest, with the adequate internal reliability ($\alpha = .915$).

With regard to the strategy use questionnaire, with Wu and Tu's (2005) suggestions being followed, the item-total correlation and the extreme group method were utilized for item analysis. Pearson product-moment correlations were conducted for the item-total correlation. Items with item-total correlations being .300 or above were retained. With reference to the extreme group method, an item was accepted if it could discriminate well between the total scores of the upper 33 percent and the lower 33 percent of the participants. An independent samples t-test was conducted for this analysis. The strategy use questionnaire ended up with seventy-two strategy items, with the satisfactory internal reliability ($\alpha = .953$).

Then, I conducted an exploratory factor analysis to extract the constructs underlying the English language knowledge test, the reading and test-taking strategy use questionnaire and the multiple-choice reading comprehension test. The results supported the presence of the construct validity for these measuring instruments.

I also applied structural equation modeling to analyze the relationship among students' English language knowledge, reading and test-taking strategy use, and their reading comprehension test performance. The result indicated that both students' English language knowledge and reading and test-taking strategy use yielded effects on their reading test performance. However, compared with strategy deployment, students' English language knowledge exercised more influences on their reading test performance. While all students' English language knowledge exerted a positive effect on their reading test performance, some of students' strategy use had an adverse impact on their reading test performance. Finally, the relationship between English language knowledge and strategy use was interactive.

To conclude, this pilot study not merely functioned to reduce test items and strategy items but also provided a general, preliminary picture of the relationship among Taiwanese senior high school students' English language knowledge, strategy use, and their reading test performance. In addition, with this pilot study, a possibility was given of analyzing the relationship among students' English language knowledge, strategy use, and their reading test performance with the use of the SEM approach. Finally, from this pilot study, I learned how to conduct exploratory factor analysis appropriately and how to construct a full latent variable model pertinent to the relation among variables of interest, with the application of SEM.

3.9 Conclusion

This chapter is concerned with the methodology of the current study. More specifically, in this chapter I discuss survey research, as well as my research design. In addition, I depict the nature of measurement, participants, instruments for gathering data, data collection procedures, methods for analyzing data and the pilot study. In the next chapter, I will address how the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance is constructed in a single modeling framework with the application of SEM and its results.

CHAPTER FOUR
MODELING THE RELATIONSHIP AMONG EFL STUDENTS' ENGLISH LANGUAGE KNOWLEDGE, STRATEGY USE, AND THEIR REARIDNG TEST PERFORMANCE: RESULTS

4.1 Introduction

This chapter is concerned with how the model regarding students' English language knowledge, reading and test-taking strategy use and their multiple-choice reading comprehension test performance is formulated, and the results of its analysis. At first, the measurement models were constructed for English language knowledge, reading and test-taking strategy use and multiple-choice reading comprehension test performance according to the results of a series of exploratory factor analyses. Then, the relationship amongst these three measurement models was examined in a single modeling framework. In other words, the relationship among students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance was hypothesized and tested by applying structural equation modeling (SEM).

This chapter is structured in the following order. First of all, I describe how the measurement models were constructed and the results. Then, I discuss how the relationship among English language knowledge, strategy use and reading test performance was formulated. The accepted model is also discussed briefly.

4.2 Constructing the measurement models

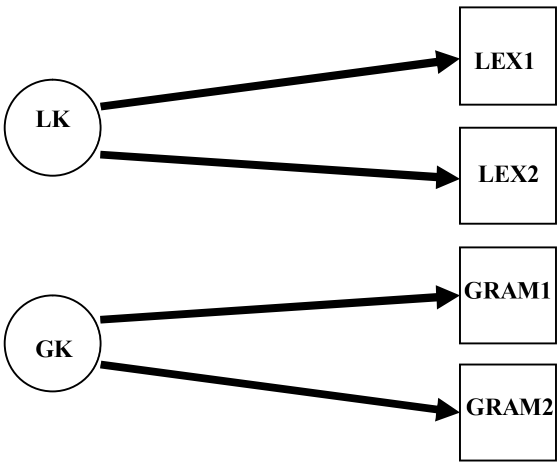
In order to formulate the measurement models for English language knowledge, strategy use, and reading test performance, I carried out an array of exploratory factor analyses (EFAs) to extract the components underlying the measuring instruments (i.e., an English language knowledge test, a strategy use questionnaire and a reading test). With the results of EFAs, I formulated the measurement models for English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance, which is addressed as follows.

4.2.1 Constructing the measurement model for English language knowledge

The result of exploratory factor analyses showed that two components were extracted from the English language knowledge test: *grammatical knowledge* (*GK*) and *lexical knowledge* (*LK*) (see pp. 249-251 for details). These two components were treated as latent variables in the present study.

Next, I further categorized the test items measuring lexical knowledge and grammatical knowledge into two subgroups based on (a) the type of test items, (b) the similar number of test items in each subgroup, and (c) the result of the reliability estimate for each subgroup. More specifically, test items with the same type were basically grouped together. The reliability estimate for each subgroup should be above .500 and items should not reduce the reliability estimate of the subgroup to which they belong.

Finally, test items assessing lexical knowledge (LK) were divided into *LEX1* and *LEX2*. These two test-item subgroups functioned as observed variables for LK (a latent variable). Test items gauging grammatical knowledge (GK) were also classified into *GRAM1* and *GRAM2*. These two test-item subgroups served as observed variables for GK (a latent variable). The constructed measurement model for English language knowledge is shown in Figure 4.1.



A circle represents a latent variable, while a rectangle represents an observed variable. → =Observed variables load on latent variables. LK=Lexical knowledge; GK=Grammatical knowledge. LEX1 consists of ten test items of the vocabulary subtest; LEX2 eleven test items of the vocabulary subtest; GRAM1 eight test items of the grammar subtest; GRAM2 eight test items of the grammar subtest.

Figure 4.1 The constructed measurement model for English language knowledge

4.2.2 Constructing the measurement model for reading and test-taking strategy use

As for the strategy use questionnaire, the result of exploratory factor analyses indicated that four components were extracted (see pp. 254-256 for details). They consisted of (a) the *monitoring, directing attention and managing the test* (*MDAMT*) process; (b) the *constructing the meaning and evaluating* (*CME*) process; (c) the *monitoring and utilizing test questions* (*MUTQ*) process; and (d) the *evaluating and marking* (*EM*) process. In the present study, these four components, serving as latent variables, were defined as strategy use processes which at a higher level than strategies are principally characterized as individuals' states of mental activity occurring during the reading test.

An examination of the *monitoring, directing attention and managing the test* process shows that this process is related to reading and question-answering processes; monitoring plays a substantial and significant role in the entire reading comprehension test. Monitoring functions to check one's comprehension of what has been processed or the current task faced (e.g., *during the reading process, I was aware that I did not understand the meaning of a word*). Then, repeating strategies can be deployed to work on incomprehensible parts (e.g., *when I did not understand the meaning of a sentence, I tried to reread it*). In addition, retrieving-linking strategies or managing-the-test strategies are employed in order to reach a possible answer or better test performance (e.g., *when I answered test questions, I tried to recall a part of the passag*e or *when I answered test questions, I tried to spend more time on difficult test questions*).

The *constructing the meaning and evaluating* process focuses on the reading process. Local or global reading strategies are manipulated to get a grip on what has been read (e.g., *during the reading process, I tried to use my words to interpret the meaning of the sentence*). Additionally, evaluation is present within this strategy use process (e.g., *when I read the passage, I tried to identify the important or less important parts of the passage*).

With respect to the *monitoring and utilizing test questions* process, the question-answering orientation is obvious with the presence of evaluating and monitoring components (e.g., *when I took the test, I tried to use clues from test questions to decide whether to read a particular part of the passage* or *when I read a sentence, I noticed it was related to test questions*).

Finally, within the *evaluating and marking* process, marking strategies with the involvement of assessment are tapped into during the entire reading comprehension test (e.g., *when I read the passage, I tried to mark key points in the passage*).

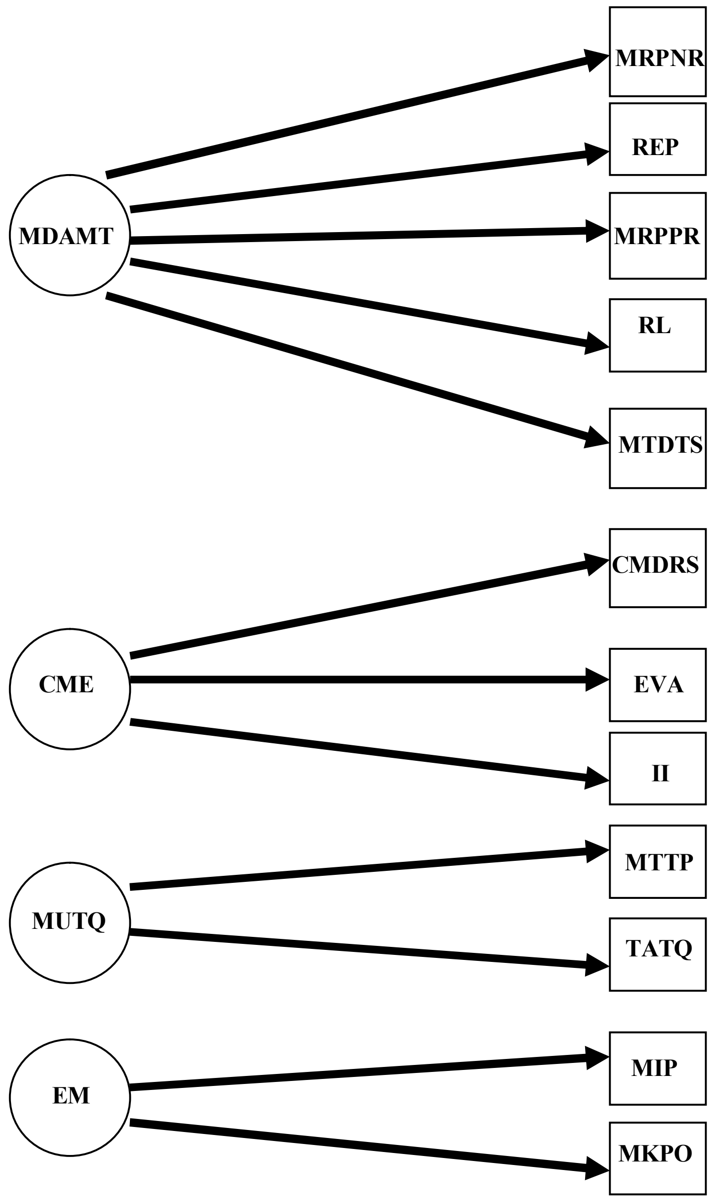
Then, I further classified strategy items in each process into subgroups, based on interpretability and the result of reliability estimates. More specifically, strategy items grouped together share similar attributes with each other and a label can be given. The reliability estimate should exceed .500 and strategy items should not decrease the reliability estimate of the subgroup to which they belong.

Finally, strategy items in the *monitoring, directing attention and managing the test* (MDAMT) process were divided into five strategy subgroups: *monitoring the reading process with negative results* (*MRPNP*), *repeating* (*REP*), *monitoring the reading process with positive results* (*MRPPR*), *retrieving-linking* (*RL*) and *managing the test with the deployment of test-taking strategies* (*MTDTS*). These strategy subgroups functioned as observed variables for the MDAMT process (a latent variable).

Similarly, strategy items included in the *constructing the meaning and evaluating* (CME) process were classified into three strategy subgroups: *constructing the meaning with the deployment of reading strategies* (*CMDRS*), *evaluating* (*EVA*) and *interacting with the input* (*II*). These three strategy subgroups served as observed variables for the CME process (a latent variable).

Further, strategy items covered by the *monitoring and utilizing test questions* (MUTQ) process were categorized into two subgroups: *monitoring the test-taking process* (*MTTP*) and *taking advantage of test questions* (*TATQ*). Both the MTTP and the TATQ subgroups functioned as observed variables for the MUTQ process (a latent variable).

Additionally, strategy items within the *evaluating and marking* (*EM*) process were split into two strategy subgroups: *marking incomprehensible parts* (*MIP*) and *marking key points or options* (*MKPO*). These two strategy subgroups served as observed variables for the EM process (a latent variable). Figure 4.2 presents the constructed measurement model for reading and test-taking strategy use.



A circle represents a latent variable, while a rectangle represents an observed variable. ➡ =Observed variables load on latent variables. MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ= Monitoring and utilizing test questions; EM=Evaluating and marking; MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA= Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP=Marking incomprehensible parts; MKPO=Marking key points or options.

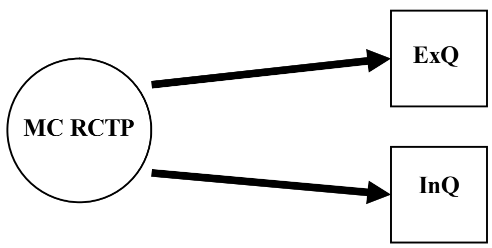
Figure 4.2 The constructed measurement model for reading and test-taking strategy use

4.2.3 Constructing the measurement model for multiple-choice reading comprehension test performance

With regard to the multiple-choice reading comprehension test, the result of exploratory factor analyses revealed that two components were extracted: *explicit questions* (*ExQ*) and *inferential questions* (*InQ*) (see p. 263 for details).

Explicit questions assessed participants' ability to read reading passages for facts, details or explicit main ideas. With little inference-drawing, participants could arrive at an answer directly after getting a general grip on part of the passage or the whole passage. Inferential questions measured participants' ability to read reading passages for implicit main ideas and to infer meanings from reading passages. Participants were required to reason the meaning for what had been read, or infer main ideas (implicit) or true statements (implicit) against the text.

Originally, I treated *explicit questions* (ExQ) and *inferential questions* (InQ) as latent variables and further categorized test items in ExQ and InQ respectively into two subgroups as observed variables. However, the result of the reliability estimate was unsatisfactory (α < .500) for three subgroups out of the four. As a result, I did not divide the test items in ExQ and InQ into two subgroups. I viewed ExQ and InQ as observed variables for multiple-choice reading comprehension test performance (MC RCTP) – a latent variable. Figure 4.3 provides the constructed measurement model for multiple-choice reading comprehension test performance.



A circle represents a latent variable, while a rectangle represents an observed variable. → =Observed variables load on latent variables. MC RCTP=Multiple-choice reading comprehension test performance; ExQ=Explicit questions; InQ=Inferential questions.

Figure 4.3 The constructed measurement model for multiple-choice reading comprehension test performance

The following section will address how the full latent variable model pertaining to the relation among English language knowledge, strategy use and reading test performance is modeled and its result.

4.3 Constructing and testing the full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance

Previous to proceeding to model the relation among students' English language knowledge, strategy use and their reading test performance, I first examined the component structure of English language knowledge (ELK) and that of reading and test-taking strategy use (RTSU) by conducting confirmatory factory analysis (CFA) with the use of the structural equation modeling (SEM) procedures. This is in order to understand the relationship between latent variables and observed variables of the measurement models of ELK and RTSU, and further to test the appropriateness of these two measurement models produced in Sections 4.2.1 and 4.2.2. I did not perform CFA to inspect the component structure of multiple-choice reading comprehension test performance, given the limited number of observed variables in this measurement model and then failure to carry out SEM for it. The results indicated that observed variables well represented their latent variables in the measurement models of ELK and RTSU. Such results provided evidence for the appropriateness of these two measurement models (see Appendices 9 and 10 for details).

After the component structures of ELK and of RTSU were validated, I carried out SEM to formulate and test the full latent variable model regarding the relationship amongst students' English language knowledge, strategy use, and their reading test performance based on theoretical underpinnings and previous studies. I adopted the model generating procedure for this SEM analysis. In other words, the initial tentative model was proposed and this hypothesized model was evaluated, based on whether this model described the collected data satisfactorily. Post-hoc adjustments were made, if necessary, to produce a model with appropriate goodness-of-fit and interpretability.

4.3.1 The hypothesized model: Model 1.1

With the constructed measurement models of English language knowledge, reading and test-taking strategy use and multiple-choice reading comprehension test performance, I specified a full latent variable model of the relationship among students' English language knowledge, strategy use, and their reading test performance. I made several hypotheses for the full latent variable model.

First of all, English language knowledge was hypothesized to have a direct influence on multiple-choice reading comprehension test performance based on three aspects. The first one is Bachman's (1990) factors that affect test scores – communicative language ability, consisting of language competence, influences test results. The second one is Bachman and Palmer's (1996) model of language ability in language use and language test performance – language knowledge exercises an influence on test performance. The third one is the findings of previous L2 reading research works (e.g., Barnett, 1986; Bernhardt & Kamil, 1995; Droop & Verhoeven, 2003; Lee & Schallert, 1997; Nassaji, 2003b; Purpura, 1997; 1998b; 1999; Shiotsu & Weir, 2007; Taillefer, 1996; Usó-Juan, 2006).

Further, reading and test-taking strategy use was postulated to exert a direct effect on multiple-choice reading comprehension test performance grounded on four aspects. The first one is Bachman's (1990) factors that affect test scores – communicative language ability subsuming strategic competence and personal attributes comprising strategy use impact upon test results. The second one is Bachman and Palmer's (1996) model of language ability in language use and language test performance – strategic competence and personal characteristics (i.e., strategy use in the current study) have an impact on test performance. The third one is the findings and the implications of previous L2 reading strategy research works (e.g., Block, 1986; 1992; Mokhtari & Reichard, 2004; Oxford *et al.*, 2004; Sarig, 1987; Sheorey & Mokhtari, 2001; Yang, 2006). The last one is the findings and the implications of previous language testing studies (e.g., Anderson, 1991; Anderson *et al.*, 1991; Cohen & Upton, 2006; 2007; Nevo, 1989; Nikolov, 2006; Phakiti, 2003; 2008; Purpura, 1997; 1998b; 1999).

In addition, lexical knowledge and grammatical knowledge were posited to be correlated with each other, predicated on the findings or the implications of previous L2 reading studies (e.g., Barnett, 1986; Shiotsu & Weir, 2007). Lexical knowledge and grammatical knowledge were also respectively postulated to be related to reading and test-taking strategy use, based on the implications of strategy related studies (e.g., Bialystok, 1981; Green & Oxford, 1995, Griffiths, 2003).

Moreover, the error terms of observed variables were hypothesized to be unrelated to one another. Finally, a disturbance (labeled as D) covering other components that influenced reading test performance but not being investigated in the current study was posited to have an effect on multiple-choice reading comprehension test performance (see Appendix 11 for the initial-hypothesized model). This model is concerned with the following research questions:

1. What is the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance?
   1.1 Do students' English language knowledge and reading and test-taking strategy use contribute to their multiple-choice reading comprehension test performance? If yes, what are their relative contributions to multiple-choice reading comprehension test performance?
   1.2 Is there a language threshold for students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance?
   1.3 What is the relationship between students' English language knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test?

Prior to testing this hypothesized model, I inspected the z-scores of the each variable to identify the possible *outliers* – in the current study, the case with the absolute value of the z-score greater than 3.000 is treated as a possible outlier (i.e., values extremely higher or lower than the other values within the data set). Thirty-three cases were pinpointed and they were dropped. Moreover, I examined the multivariate normality of the data set. Kline (1998) suggests that when the absolute value of kurtosis is over 10, a distribution of the data is regarded as a non-normal distribution. In the current study, this criterion was adopted. The result of the assessment of multivariate normality indicated that the multivariate kurtosis was above the acceptable limits (18.840 > 10), suggesting the obvious multivariate non-normality of the data. According to the result of the Mahalanobis-d-squared, I removed thirty-one cases. The multivariate kurtosis value reduced to 9.752, which was within the accepted limits (< 10). Then, I performed SEM to test this hypothesized full latent variable model.

4.3.2 The results for Model 1.1

The result for Model 1.1 indicated that the chi-square statistic of 219.093 reached statistical significance at the .050 level. The values of the GFI, AGFI, CFI and TLI respectively were .968, .951, .979 and .972, all above the cut-off value. Furthermore, the RMSEA of .036 was below the threshold level (< .060). Based on the results, this hypothesized model seemed to depict the gathered data fairly and should be accepted. However, an inspection of the parameter estimation revealed that the effects of the *constructing the meaning and evaluating* process, and the *evaluating and marking* process on multiple-choice reading comprehension test performance were nonsignificant at the 5% level. Additionally, the relationship between the *monitoring and utilizing test questions* process and either grammatical knowledge or lexical knowledge did not arrive at statistical significance. Furthermore, the variance estimate of D1 was also nonsignificant at the 5% level. As a consequence, this model was respecified.

4.3.3 The hypothesized model: Model 1.2

According to the results of Model 1.1, I made several post-hoc adjustments to respecify the model grounded on previous studies, modification indices and interpretability. To illustrate, grammatical knowledge was postulated to have an impact on lexical knowledge in this reading test, predicated on the implications given in Nassaji's (2003a) and Paribakht's (2004) studies.

In addition, a strategy use process was hypothesized to display an effect on other strategy use processes based on two aspects. The first one is Bachman and Palmer's (1996) model of language ability in language use and language test performance – strategic competence (i.e., metacognitive strategies) shows an influence on personal attributes (i.e., strategy use in the current study). The second one is the findings in language testing studies concerning strategy use (e.g., Phakiti, 2008; Purpura, 1997; 1998b; 1999).

Further, on the one hand, English language knowledge was hypothesized to show an effect on strategy deployment predicated on three aspects. The first one is Bachman's (1990) model of communicative language ability and Bachman and Palmer's (1996) model of language ability in language use and language test performance – language knowledge influences strategic competence. The second one is the implications offered in previous L2 reading strategy studies (e.g., McLeod & McLaughlin, 1986; Oxford *et al.*, 2004; Stevenson *et al.*, 2003; Upton & Lee-Thompson, 2001). The third one is the implications provided in previous L1-L2 reading studies (e.g., Bossers, 1991; Clarke, 1980; Lee & Schallert, 1997; Taillefer, 1996; Walter, 2004; Yamashita, 2002).

On the other hand, strategy employment was postulated to have an impact on English language knowledge, grounded on two aspects. The first one is Bachman and Palmer's (1996) model of language ability in language use and language test performance – strategic competence affects language knowledge. The second one is the findings or the implications provided in previous strategy research works (e.g., Fraser, 1999; Kern 1989; Purpura, 1997; 1998b; 1999).

Finally, the uniquenesses (errors) related to observed variables were posited to have a relationship with one another. Such a hypothesis was grounded on whether the result was interpretable, exacerbated the overall model fit or reached statistical significance.

While I hypothesized the relationship between language knowledge and strategy use predicated on the implications of Bachman's (1990) and Bachman and Palmer's (1996) models of language ability and previous related research works, I had to acknowledge that the causal effect between strategy deployment and language knowledge was determined by modification indices and model fit statistics. I did so based on the following reason – the causal direction between strategy deployment and language knowledge has not been definitely decided yet. Then, according to the implications of Bachman and Palmer's (1996) model of language ability and previous related studies, I should have hypothesized that all strategy use processes had an effect on both types of language knowledge and vice versa. However, doing so would complicate the entire model and make the model difficult to be interpreted. Therefore, I finally allowed the causal effect paths between all strategy use processes and both types of language knowledge to be determined by modification indices and model fit statistics.

To sum up, in the current study SEM was utilized in an exploratory manner in the following aspects (a) what type of English language knowledge has an effect on what type of strategy use; (b) what type of strategy use has an effect on what type of English language knowledge; (c) a type of strategy use process has an effect on another type of strategy use process (d) an observed variable (i.e., a strategy subgroup) has an effect on a latent variable (i.e., multiple-choice reading comprehension test performance); (e) error-correlations.

4.3.4 The results for Model 1.2

Based on the abovementioned adjustments, I tested thirty-seven SEM models and inspected their model fit indices. Finally, a model with appropriate good-of-fit and interpretability was produced (see Appendix 11 for the final accepted model). In this section, the model fit statistics of this model are concentrated on to justify the appropriateness for accepting the model. The model fit indices are listed in Table 4.1.

Table 4.1 The model fit indices for the full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the entire group

Model fit indices	Levels of acceptable fit	Evaluation results
χ^2	Nonsignificant with the p-value > .050	Good (110.776 with p = .461 > .050)
GFI	> .900	Very good (GFI = .984)
AGFI	> .900	Very good (AGFI = .975)
CFI	> .950	Very good (CFI = 1.000)
TLI	> .950	Very good (TLI = 1.000)
RMSEA	< .060	Very good (RMSEA = .003)

Note. N=770. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

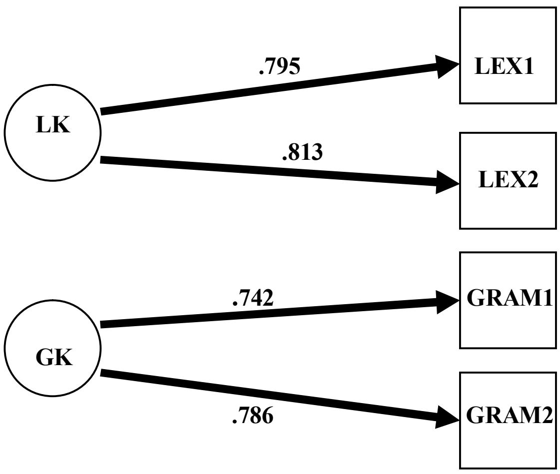
As shown in the above table, the chi-square statistic of 110.776 (much smaller than the previous one) was nonsignificant at the .050 level: its p-value was .461. The values of the GFI, the AGFI, the CFI and the TLI corresponded to .984 .975, 1.000 and 1.000 respectively, all of which exceeded the cut-off value. Similarly, the RMSEA result of .003 was well below the .060 threshold. In brief, the aforementioned model fit indices suggested that Model 1.2 was a fair representation of the sample data and provided strong evidence for the acceptance of this model. All the effect paths and correlations listed in the model were statistically significant at the .050 level ($p < .050$). Further, I also performed the bootstrap analysis to examine whether indirect effects revealed in this model were statistically significant. The result of the bootstrap analysis showed that all the indirect effects reached statistical significance, except the indirect effects of the *monitoring and utilizing test questions* (MUTQ) process on explicit questions (ExQ) and inferential questions (InQ).

In the following sections, I will first examine the individual measurement models and then shift to the structural model depicting the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance. The following criteria are adopted to describe the effect strength. Firstly, the effect strength below .100 is regarded as a *trivial* effect. Secondly, the effect strength ranging from .100 to .299 is viewed as a *weak* effect. Thirdly, the effect strength varying from .300 to .599 is treated as a *moderate* effect. Finally, the effect strength .600 or above is thought of as a *strong* effect.

4.3.4.1 The measurement model of English language knowledge

In this section, I first briefly depict what the measurement model of English language knowledge encompasses. Then, I examine factor loadings shown in this measurement model.

Figure 4.4 illustrates that English language knowledge is symbolized by two components: lexical knowledge (LK) and grammatical knowledge (GK). Both LK and GK serve as latent variables, each followed by two indicators – LEX1 and LEX2 for LK; GRAM1 and GRAM 2 for GK. Each indicator subsumes several lexical test items or grammatical test items.



LK=Lexical knowledge; GK=Grammatical knowledge. LEX1 consists of ten test items of the vocabulary subtest; LEX2 eleven test items of the vocabulary subtest; GRAM1 eight test items of the grammar subtest; GRAM2 eight test items of the grammar subtest.
➡ =Observed variables load on latent variables.

Figure 4.4 The measurement model of English language knowledge for the entire group

As expected, within this measurement model, the LEX1 and the LEX2 test-item subgroups well account for lexical knowledge (LK), with loadings of .795 and .813 respectively. The GRAM1 and the GRAM2 test-item subgroups also adequately explain grammatical knowledge (GK), with loadings of .742 and .786 respectively. These results suggest fair relationships between latent and observed variables in the measurement model.

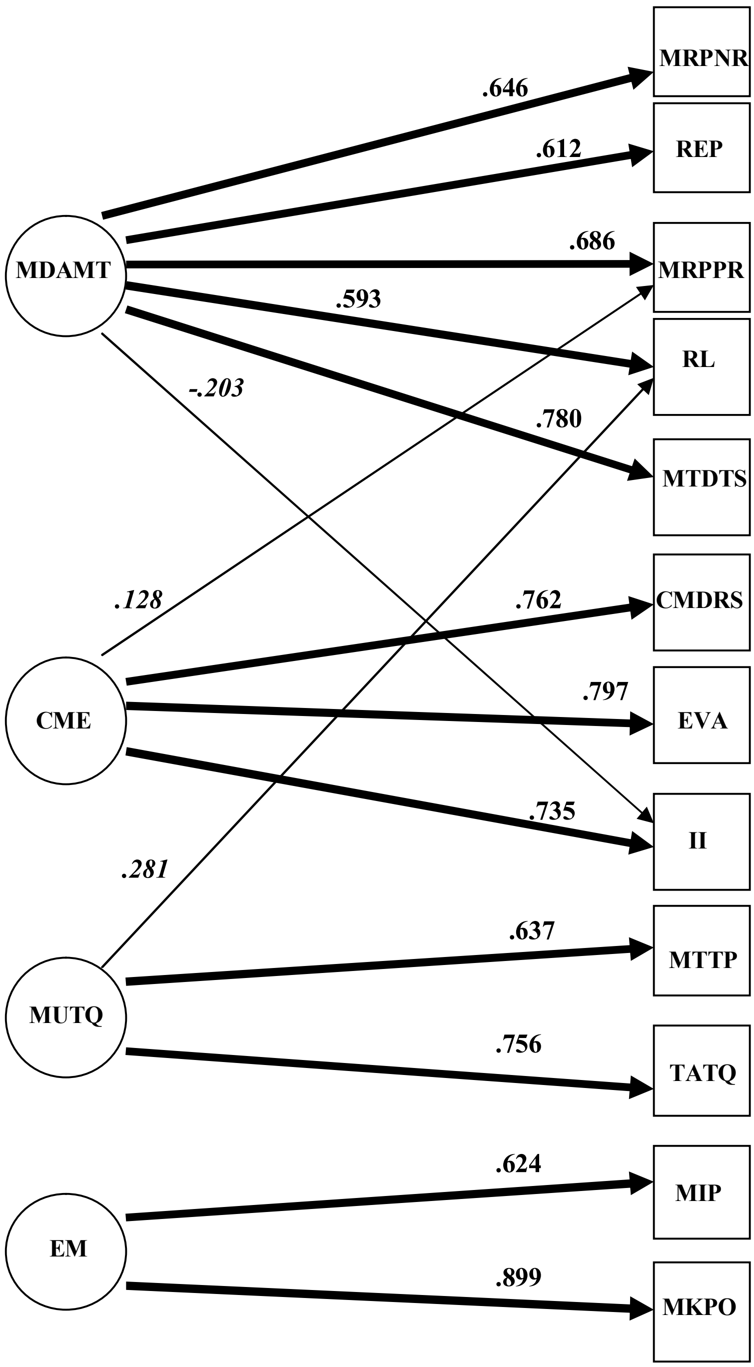
To summarize, within this measurement model, the LEX1 and the LEX2 test-item subgroups well function as indicators for lexical knowledge and so do the GRAM1 and the GRAM2 test-item subgroups for grammatical knowledge. English language knowledge is not a single-facet construct, which at least consists of lexical knowledge and grammatical knowledge.

4.3.4.2 The measurement model of reading and test-taking strategy use

In this section, I first briefly describe what the measurement model of reading and test-taking strategy use subsumes. Next, I inspect factor loadings manifested in this measurement model.

As indicated in Figure 4.5, reading and test-taking strategy use (RTSU) is characterized by four strategy use processes: (a) the *monitoring, directing attention and managing the test* (MDAMT) process; (b) the *constructing the meaning and evaluating* (CME) process; (c) the *monitoring and utilizing test questions* (MUTQ) process; and (d) the *evaluating and marking* (EM) process. These processes are latent variables, followed by two to five observed variables.

Within this measurement model, the *monitoring the reading process with negative results* (MRPNR), the *repeating* (REP), the *monitoring the reading process with positive results* (MRPPR), the *retrieving-linking* (RL) and the *managing the test with the deployment of test-taking strategies* (MTDTS) strategy subgroups well explain the *monitoring, directing attention and managing the test* (MDAMT) process. The factor loadings correspond to .646, .612, .686, .593 and .780. Interestingly, among these strategy subgroups, the *managing the test with the deployment of test-taking strategies* (MTDTS) strategy subgroup measures the MDAMT process best. The result indicates that this strategy subgroup is beneficial most to the *monitoring, directing attention and managing the test* process in this reading test-taking process, suggesting students' strong test-taking tendency towards this reading test.



MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; EM=Evaluating and marking; MRPNR=Monitoring the reading process with negative results; REP= Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA= Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP= Marking incomprehensible parts; MKPO=Marking key points or options. ➡ =Observed variables load on latent variables. → =Observed variables cross-load on latent variables.

Figure 4.5 The measurement model of reading and test-taking strategy use for the entire group

Turning to the *constructing the meaning and evaluating* (CME) process, this strategy use process is well explained by the *constructing the meaning with the deployment of reading strategies* (CMDRS), the *evaluating* (EVA) and the *interacting with the input* (II) strategy subgroups, with loadings of .762, .797 and .735. A similarity in loadings implies that these strategy subgroups make similar contributions to the *constructing the meaning and evaluating* process in the course of this reading comprehension test. When sitting this reading test, students invoke local and global reading strategies to equally facilitate their having a grip on the input, make an appropriate judgment if needed, and interact and communicate with what they process.

Similarly, the *monitoring and utilizing test questions* (MUTQ) process, is properly accounted for by its indicators, the *monitoring the test-taking process* (MTTP) and the *taking advantage of test questions* (TATQ) strategy subgroups, with loadings of .637 and .756. The result suggests that these EFL students, as taking this multiple-choice reading test, rest on monitoring strategies and strategies capitalizing on test questions to assist in their supervising their overall test-taking process and enhancing test performance.

Finally, within the *evaluating and marking* (EM) process, the loadings that the *marking incomprehensible parts* (MIP) and the *marking key points or options* (MKPO) strategy subgroups produce respectively correspond to .624 and .899. Such a result reveals that these marking strategy subgroups well explain the EM process. Intriguingly, the *marking key points or options* strategy subgroup better accounts for the EM process. This indicates that the *marking key points or options* strategy subgroup is more profitable to the *evaluating and marking* process than the *marking incomprehensible parts* strategy subgroup. The finding makes sense, as one can imagine that these students' assessing what have been processed and conducting marking on those related to answering test questions contribute to their reading test performance more directly than their marking incomprehensible portions.

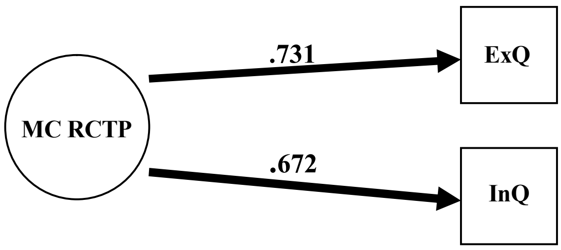
The model presented in Figure 4.5 also manifests that three strategy subgroups assess more than one latent variable. These cross-loadings suggest that these strategy subgroups do not have a unique linkage with one strategy use process. Among these cross-loadings, the *interacting with the input* (II) strategy subgroup shows a positive relationship with the *constructing the meaning and evaluating* (CME) process (.735) but a negative one with the *monitoring, directing attention and managing the test* (MDAMT) process (-.203). Such an interesting result indicates that the *interacting with the input* strategy subgroup serves as a beneficial strategy subgroup in the *constructing the* *meaning and evaluating* process while a detrimental strategy subgroup in the *monitoring, directing attention and managing the test* process. It also supports the notion that students need to employ strategies on appropriate occasions.

To summarize, strategy subgroups serve properly as indicators for strategy use processes in this measurement model despite the presence of three cross-loadings. Strategy deployment, rather than a single-facet construct, is a multi-facet construct, as Purpura's (1997; 1999) studies have demonstrated.

4.3.4.3 The measurement model of multiple-choice reading comprehension test performance

In this section, I first outline what the measurement model of multiple-choice reading comprehension test performance comprises. Then, I examine factor loadings revealed in this measurement model.

As can be seen in Figure 4.6, multiple-choice reading comprehension test performance is simply represented by one component: multiple-choice reading comprehension test performance (MC RCTP). MC RCTP, a latent variable, is assessed by two observed variables: explicit questions (ExQ) and inferential questions (InQ).



MC RCTP=Multiple-choice reading comprehension test performance; ExQ=Explicit questions; InQ=Inferential questions.
→ =Observed variables load on latent variables.

Figure 4.6 The measurement model of multiple-choice reading comprehension test performance for the entire group

In this measurement model, both explicit questions and inferential questions explain multiple-choice reading comprehension test performance well, with loadings of .731 and .672 respectively. The result gives an implication. To explain, answering some explicit questions simply entails students' reading and comprehending part of the passage or processing the passage at the lexical or the syntactic levels, which triggers more the bottom-up reading. On the other hand, answering inferential questions necessitates students' piecing together information that may spread across the passage, and then drawing inferences after digesting the input, which elicits more the top-down or the interactive reading. Given what is mentioned, these intermediate-beginning or intermediate EFL students, when sitting this reading comprehension test, appear to conduct the bottom-up, the top-down or the interactive reading to process the passages to a similar degree. Such an implication contradicts what a number of L2 studies suggest (e.g., Barnett, 1989; Carrell, 1988; Clarke, 1980; Purpura 1997; 1999) – L2 learners' tendency towards the bottom-up processing in L2 reading.

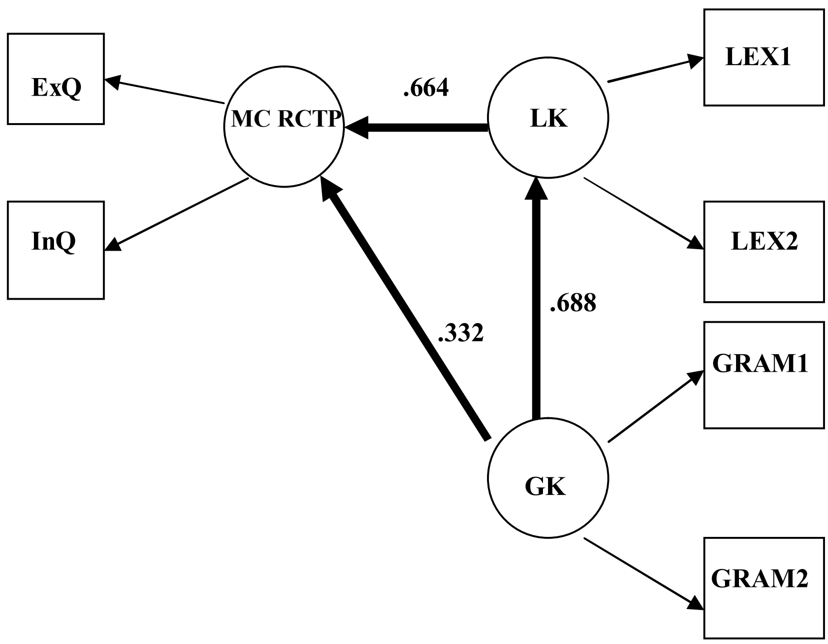
In summary, both explicit and inferential questions measure appropriately multiple-choice reading comprehension test performance within this measurement model. The following sections will concentrate on the structural model pertinent to the relation among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance.

4.3.4.4 The relationship between English language knowledge and multiple-choice reading comprehension test performance

In this section, I first examine an effect that lexical knowledge exerts on multiple-choice reading comprehension test performance and then turn to an impact that grammatical knowledge has on multiple-choice reading comprehension test performance.

As for the relation between English language knowledge and reading test performance, I had hypothesized that both lexical knowledge and grammatical knowledge would exercise influences on multiple-choice reading comprehension test performance. Such hypotheses were accepted and these causal effect paths were significantly observed in the full latent variable model.

As presented in Figure 4.7, expectedly, lexical knowledge (LK) shows a strong, direct, positive impact on multiple-choice reading comprehension test performance (MC RCTP), with a value of .664. The result suggests that lexical knowledge directly contributes to multiple-choice reading comprehension test performance greatly. An important role that LK plays in this reading comprehension test performance is illustrated.



MC RCTP=Multiple-choice reading comprehension test performance; ExQ=Explicit questions; InQ=Inferential questions; LK= Lexical knowledge; GK=Grammatical knowledge. ➡ =A latent variable has an effect on another latent variable.

Figure 4.7 The relationship between English language knowledge and multiple-choice reading comprehension test performance for the entire group

Distinct from lexical knowledge (LK), grammatical knowledge (GK) displays a moderate, direct, positive effect on multiple-choice reading comprehension test performance (MC RCTP), with a value of .332. In terms of the direct effect, LK carries more weight than GK in this L2 reading test performance. However, there is still something else. Surprisingly, by means of LK, GK also has a significant, moderate, indirect, positive impact on MC RCTP, with a value of .457. Such a result suggests that perhaps students' getting access to grammatical knowledge aids them in the vocabulary inferencing processing to figure out the meanings of unfamiliar words, thereby contributing to their reading test performance, as implied in Nassaji's (2003a) and Paribakht's (2004) research works.

In summary, students' lexical knowledge has a *strong*, direct, positive effect and grammatical knowledge a *moderate* impact on their multiple-choice reading comprehension test performance. Additionally, their grammatical knowledge, by means of lexical knowledge, yields a moderate, *indirect*, positive effect on their multiple-choice reading comprehension test performance.

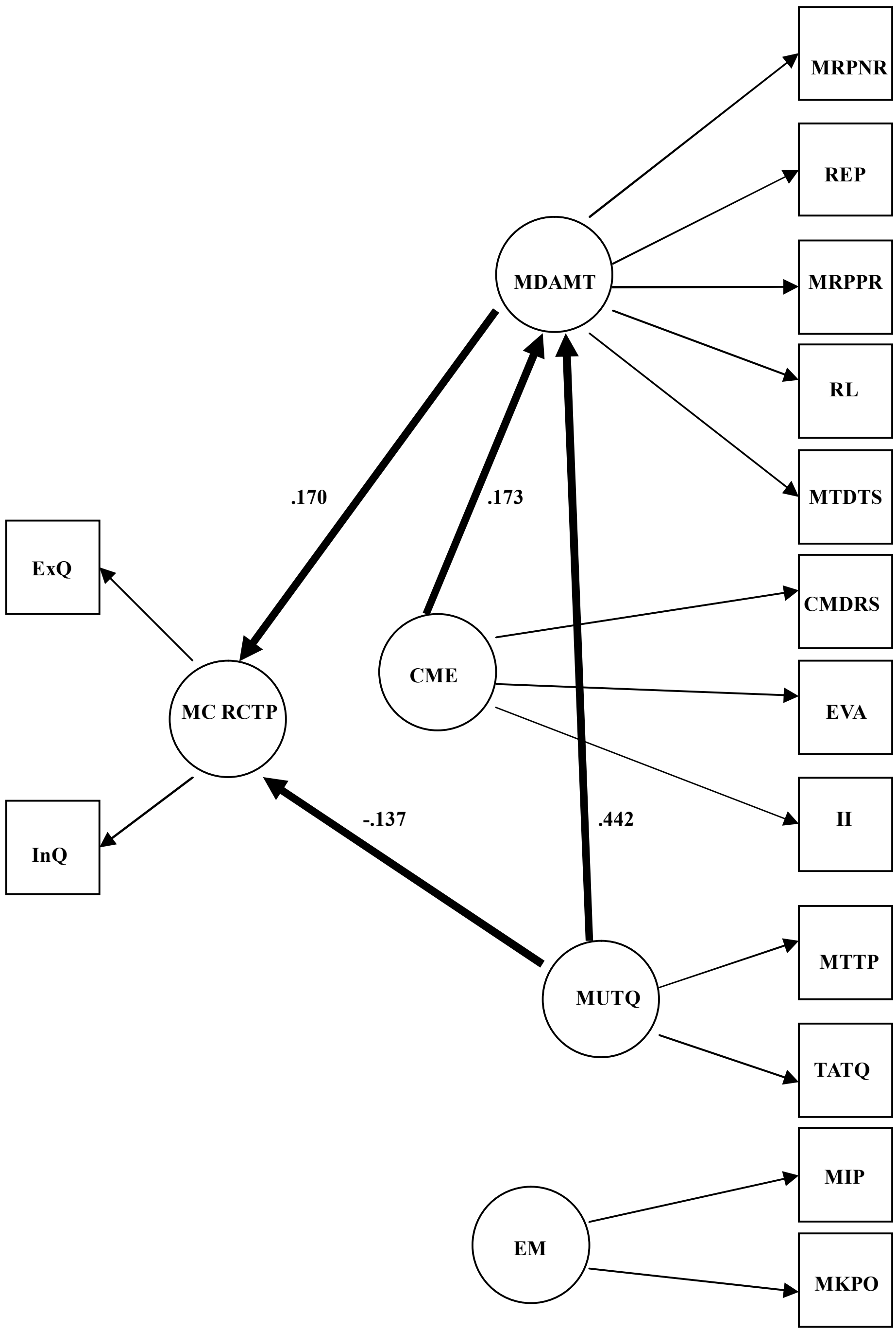
4.3.4.5 The relationship between reading and test-taking strategy use and multiple-choice reading comprehension test performance

In this section, I inspect an effect that strategy use processes yield on multiple-choice reading comprehension test performance.

Originally, I had hypothesized that all strategy use processes would have an impact on multiple-choice reading comprehension test performance. However, contrary to expectations, not all hypotheses were accepted in the full latent variable model. Stated another way, not all strategy use processes displayed a significant effect on multiple-choice reading comprehension test performance.

The *evaluating and marking* (EM) process is the only strategy use process that has no significant, direct or indirect impacts on multiple-choice reading comprehension test performance (MC RCTP). The EM process is composed of marking strategies. Then, it should show an effect on reading test performance, as Sarig's study (1987) suggests. However, this is not observed in the current data set. In an effort to examine the relation between the EM process and reading test performance, a model was tested in which the *marking incomprehensible parts* (MIP) and the *marking key points or options* (MKPO) strategy subgroups, observed variables for the EM process, were hypothesized to exert an effect on MC RCTP. But the effect paths were not statistically significant. This reveals that even the marking strategy subgroups, indicators of the EM process, do not impact on MC RCTP.

The reason why the *evaluating and marking* (EM) process influencing reading test performance is not captured in the model can be attributed to what Bialystok (1981) has remarked, "time spent on some of the strategies is more profitable than time spent on some of the others" (p. 33). In this reading test-taking course, students perhaps consider it not beneficial to make a judgment about what they read and to conduct marking on what they deem is important. They prefer to invoke other strategies or turn to other cognitive resources to overcome their comprehension breakdowns or optimize their test performance. Therefore, the effect of the EM process on reading test performance is not observed. Apart from this, it can also be explained from the information-processing perspective. These students might utilize marking strategies subsumed by the EM process in automatizing and restructuring processes; consequently, the effect of this strategy use process on reading test performance fails to be manifested in the full latent variable model.



MC RCTP=Multiple-choice reading comprehension test performance; MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; EM=Evaluating and marking.
➡ =A latent variable has an effect on another latent variable.

Figure 4.8 The relationship between reading and test-taking strategy use and multiple-choice reading comprehension test performance for the entire group

Among the strategy use processes impacting on multiple-choice reading comprehension test performance (MC RCTP), how they affect MC RCTP is different. As shown in Figure 4.8, the *monitoring, directing attention and managing the test* (MDAMT) process exerts a weak, direct, positive effect on MC RCTP, with a value of .170. The result suggests that these EFL students' deployment of monitoring, repeating, retrieving-linking and managing-the-test strategies covered by this strategy use process weakly promotes their reading test performance in a direct manner.

Distinct from the *monitoring, directing attention and managing the test* process, the *constructing the meaning and evaluating* (CME) process yields no direct effect on multiple-choice reading comprehension test performance (MC RCTP). This result comes as a surprise. As implied in Pressley and Afflerbach's (1995) model of constructively responsive reading and other previous research works relevant to L2 reading strategies (e.g., Hosenfeld, 1984; Oxford *et al.*, 2004; Padron & Waxman, 1988; Yang, 2006), reading strategies regarding constructing the meaning of the text have an impact on reading comprehension. Then, the CME process comprising global and local reading strategies should display a direct effect on MC RCTP in this test-taking setting. A plausible explanation for the CME process not impacting upon MC RCTP directly rests on the fact that in this reading comprehension test, students have fewer tendencies to directly deploy strategies involved in this strategy use process. As a result, the direct effect that the CME process exerts on MC RCTP fails to be observed in the full latent variable model.

Nonetheless, by means of the *monitoring, directing attention and managing the test* (MDAMT) process, the *constructing the meaning and evaluating* (CME) process exercises a trivial, indirect, positive influence on multiple-choice reading comprehension test performance (MC RCTP), with a value of .029. The result implies that these students invoke constructing-the-meaning and evaluating strategies contained by the CME process in concert with monitoring, repeating, retrieving-linking and managing-the-test strategies subsumed by the MDAMT process to make an indirect contribution to their reading test performance in this test-taking context.

Like the *monitoring, directing attention and managing the test* process, the *monitoring and utilizing test questions* (MUTQ) process exerts a direct influence on multiple-choice reading comprehension test performance (MC RCTP). However, rather than a direct, *positive* one, the MUTQ process yields a weak, direct, *negative* effect on MC RCTP, with a value of -.137. The result suggests that students' employing monitoring and utilizing-test-questions strategies is weakly detrimental to their performance on the reading test.

Unexpectedly, apart from the direct effect, the *monitoring and utilizing test questions* (MUTQ) process, through the *monitoring, directing attention and managing the test* (MDAMT) process, also has a significant indirect impact on multiple-choice reading comprehension test performance (MC RCTP). However, instead of a negative one, it is a positive effect, with a value of .075. The result suggests that students' employment of monitoring and utilizing-test-questions strategies, like their use of constructing-the-meaning and evaluating strategies, enhances their reading test performance via their deployment of monitoring, repeating, retrieving-linking and managing-the-test strategies. The result here also indicates the complexity of the MUTQ process, given that this strategy use process can not operate in isolation if aiming to contribute to reading test performance.

In summary, not all students' strategy use processes yield an effect on their multiple-choice reading comprehension test performance in this test-taking setting. Among all strategy use processes, their *monitoring, directing attention and managing the test* process represents the only strategy use process displaying a direct effect on their multiple-choice reading comprehension test performance. Further, students' *constructing the meaning and evaluating* process, and *monitoring and utilizing test questions* process have a trivial, indirect, positive effect on their multiple-choice reading comprehension test performance by means of the *monitoring, directing attention and managing the test* process. Additionally, their *monitoring and utilizing test questions* process shows a weak, direct, negative effect on their multiple-choice reading comprehension test performance.

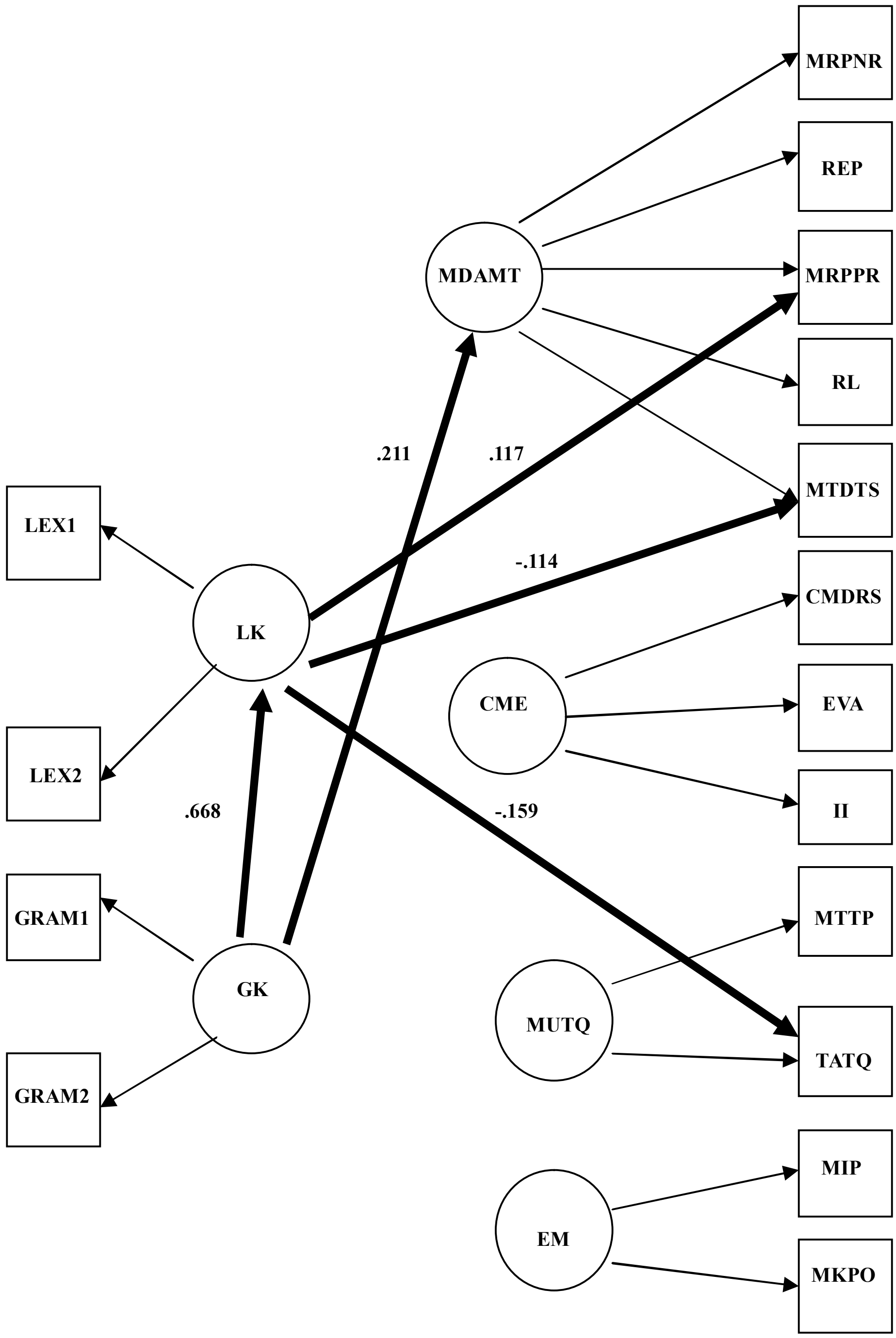
4.3.4.6 The relationship between English language knowledge and reading and test-taking strategy use

In this section, first I examine an effect that English language knowledge (i.e. lexical knowledge and grammatical knowledge) exerts on reading and test-taking strategy use (i.e., strategy use processes and strategy subgroups) and then shift to an impact that reading and test-taking strategy use has on English language knowledge.

The full latent variable model captures interactive paths between English language knowledge and strategy use. On the one hand, English language knowledge affects strategy employment. To illustrate, lexical knowledge (LK) shows a weak, direct, *positive* effect on the *monitoring the reading process with positive results* (MRPPR) strategy subgroup, with a value of .117 (see Figure 4.9). In addition, grammatical knowledge (GK) shows a weak, direct, positive impact on the *monitoring, directing attention and managing the test* (MDAMT) process, with a value of .211. Interestingly, LK has a weak, direct, *negative* effect on the *managing the test with the deployment of test-taking strategies* (MTDTS) strategy subgroup, with a value of -.114 and on the *taking advantage of test questions* (TATQ) strategy subgroup, with a value of -.159 (see Figure 4.9). Through LK, GK has a weak, indirect, negative impact on the *taking advantage of test questions* (TATQ) strategy subgroup, with a value of -.114 (see Table 4.2). Such results suggest that students' English language knowledge inhibits their deployment of managing-the-test and taking-advantage-of-test-questions strategies.

Based on the aforementioned results, two implications are provided. Firstly, lexical knowledge and grammatical knowledge does not always contribute to strategy utilization within this reading test-taking context. Secondly, deploying some strategies requires a certain amount of linguistic processing. In order to invoke these strategies adequately and effectively, students need to rest on language knowledge first, make initial sense of the input, and then assess whether to deploy strategies, what strategy to be employed and how to utilize strategies.

The results here provide empirical evidence for Bachman's (1990) model of language ability in which language knowledge impacts upon strategic competence and appear to give empirical evidence for the resolution of the direction of causality between strategy employment and language knowledge. However, the real case is not that simple, since the full latent variable model also captures the path that strategy deployment influences English language knowledge, which is addressed in the following.



LK=Lexical knowledge; GK=Grammatical knowledge; MDAMT=Monitoring, directing attention and managing the test; MRPPR= Monitoring the reading process with positive results; MTDTS=Managing the test with the deployment of test-taking strategies; TATQ=Taking advantage of test questions. ➘ =English language knowledge has an effect on strategy use.

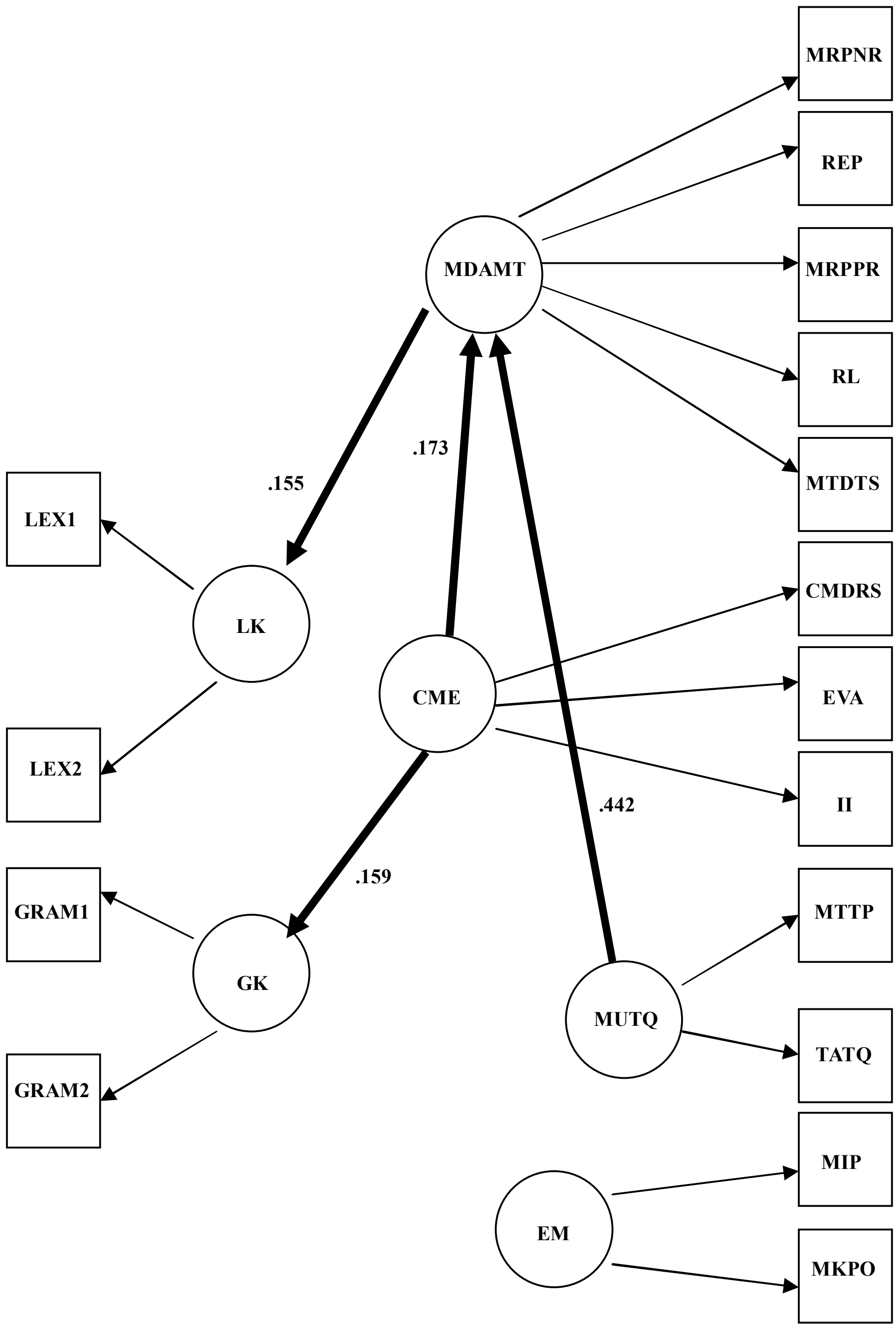
Figure 4.9 The relationship between English language knowledge and reading and test-taking strategy use for the entire group (part I)

Counter to my expectation, strategy deployment always influences English language knowledge positively (see Figure 4.10). To explain, the *monitoring, directing attention and managing the test* (MDAMT) process displays a weak, direct, positive effect on lexical knowledge (LK), with a value of .155 and the *constructing the meaning and evaluating* (CME) process on grammatical knowledge (GK), with a value of .159 (see Figure 4.10). Further, the CME process, via the MDAMT process, exerts a trivial, indirect, positive effect on LK, with a value of .027 (see Table 4.2). Following the same path, the *monitoring and utilizing test questions* (MUTQ) process also has a trivial, indirect, positive impact on LK, with a value of .069 (see Table 4.2).

The results stated above suggest that strategy use facilitates English language knowledge access or development directly or indirectly. The direction of causality between language knowledge and strategy deployment reverses what has been noted on page 111 and controverts the causal relationship between language knowledge and strategic competence, presented in Bachman's (1990) model of language ability. In brief, in this EFL reading test, some strategy use processes affect part of English language knowledge, whereas part of English language knowledge influences some strategy subgroups or strategy use processes.

A close inspection of the relationship between English language knowledge and strategy use processes shows that the *evaluating and marking* (EM) process has no impact on English language knowledge and vice versa. Students' deploying marking strategies appears to entail little linguistic processing and contribute little to their language knowledge access or development within this reading test-taking situation.

In summary, within this reading test, students' English language knowledge and strategy use interact with each other. Sometimes their English language knowledge affects strategy use and sometimes their strategy use influences English language knowledge. Interestingly, students' English language knowledge shows a positive or negative effect on strategy use whereas their strategy use always has a positive impact on English language knowledge.



LK=Lexical knowledge; GK=Grammatical knowledge; MDAMT=Monitoring, directing attention and managing the test; CME= Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions. ➘ =Strategy use has an effect on English language knowledge.

Figure 4.10 The relationship between English language knowledge and reading and test-taking strategy use for the entire group (part II)

Table 4.2 Effects of English language knowledge on reading and test-taking strategy use and effects of reading and test-taking strategy use on English language knowledge for the entire group

		Effects
Effects of ELK on RTSU	LK→MRPPR	.117
	LK→MTDTS	-.114
	LK→TATQ	-.159
	GK→MDAMT	.211
	GK→TATQ (GK→LK→ TATQ; GK→MDAMT→ LK→TATQ)	-.114
Effects of RTSU on ELK	MDAMT→LK	.155
	CME→GK	.159
	CME→MDAMT→LK	.027
	MUTQ→MDAMT→LK	.069

Note. LK=Lexical knowledge; MRPPR=Monitoring the reading process with positive results; MTDTS=Managing the test with the deployment of test-taking strategies; TATQ=Taking advantage of test questions; GK=Grammatical knowledge; MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ= Monitoring and utilizing test questions.

4.3.4.7 The relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance

In this section, I first inspect an indirect effect that multiple-choice reading comprehension test performance (MC RCTP) receives from English language knowledge (ELK) and reading and test-taking strategy use (RTSU). Next, I examine a total effect, consisting of a direct effect and an indirect effect in the entire model, which ELK and RTSU show on MC RCTP.

When looking at the relation among English language knowledge, strategy use and reading test performance, I found that there was something different. More indirect effects that reading and test-taking strategy use shows on multiple-choice reading comprehension test performance are manifested in the full latent variable model. More specifically, by means of English language knowledge, strategy use influences reading test performance in an indirect way. To illustrate, through lexical knowledge, grammatical knowledge or both, the *monitoring, directing attention and managing the test*, the *constructing the meaning and evaluating* and the *monitoring and utilizing test questions* processes have a trivial or weak, indirect, positive impact on reading test performance (see Table 4.3). These results suggest that these strategy use processes more or less make indirect contributions to students' performance on the reading comprehension test with the assistance of their lexical knowledge, grammatical knowledge, or both.

Table 4.3 Indirect effects of reading and test-taking strategy use on multiple-choice reading comprehension test performance through English language knowledge for the entire group

	Effects
MDAMT→**LK**→MC RCTP	.103
CME→MDAMT→**LK**→MC RCTP	.018
CME→**GK**→MC RCTP	.053
CME→**GK**→MDAMT→**LK**→MC RCTP	.003
CME→**GK**→**LK**→MC RCTP	.073
MUTQ→MDAMT→**LK**→MC RCTP	.045

Note. MDAMT=Monitoring, directing attention and managing the test; LK=Lexical knowledge; MC RCTP=Multiple-choice reading comprehension test performance; CME=Constructing the meaning and evaluating; GK=Grammatical knowledge; MUTQ=Monitoring and utilizing test questions.

Unexpectedly, an indirect effect that English language knowledge yields on reading test performance by means of strategy deployment is reflected in this full latent variable model. Grammatical knowledge has a trivial, indirect impact on multiple-choice reading comprehension test performance through the *monitoring, directing attention and managing the test* process, with a value of .036. The result suggests that students' grammatical knowledge makes a trivial, indirect contribution to their reading test performance through their strategy use.

When the total effects of English language knowledge and strategy use on reading test performance are examined simultaneously, a clear picture of the extent to which English language knowledge and strategy use have an effect on reading test performance is shown. As indicated in Table 4.4, within English language knowledge, lexical knowledge (LK) and grammatical knowledge (GK) respectively display a strong, positive total effect on multiple-choice reading comprehension test performance (MC RCTP), with values of .664 and .846. However, GK has more impacts on MC RCTP than LK. This suggests that students rest on their grammatical knowledge more greatly than their lexical knowledge to deal with this reading test. Such a finding can be partially explained as follows.

In the classroom, most Taiwanese English teachers at the senior high school level, as teaching reading passages, get accustomed to elucidating grammatical rules or parsing complex sentences with the use of grammatical rules to construct the meaning of these sentences. This type of the teaching style may invisibly, gradually influence students' reading behaviors. Therefore, participants, as senior high school students, consciously or subconsciously get access to grammatical knowledge heavily, when taking this reading comprehension test, to process what they read.

Table 4.4 Total effects of English language knowledge and reading and test-taking strategy use on multiple-choice reading comprehension test performance for the entire group

		Effects
ELK	LK→MC RCTP	.664
	GK→MC RCTP	.846
RTSU	MDAMT→MC RCTP	.273
	CME→MC RCTP	.182
	MUTQ→MC RCTP	-.017

Note. ELK=English language knowledge; RTSU=Reading and test-taking strategy use; GK= Grammatical knowledge; MC RCTP=Multiple-choice reading comprehension test performance; LK=Lexical knowledge; MDAMT=Monitoring, directing attention and managing the test; CME= Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions.

Inspecting the total effects of strategy use on reading test performance, I found that the effects, except that for the *monitoring and utilizing test questions* (MUTQ) process, were slightly larger than those mentioned in Section 4.3.4.5 (e.g., from .170 to .273 for the *monitoring, directing attention and managing the test* process). The finding makes sense since in this section indirect effects that strategy use exerts on reading test performance are also taken into account. Surprisingly, the effect of the *constructing the meaning and evaluating* (CME) process on reading test performance becomes even stronger (from .029 to .182). This indicates a close relation between the CME process and English language knowledge. It appears that students need to rely upon a certain amount of English language knowledge in order to effectively deploy reading and evaluating strategies included in the CME process. Despite the effects being larger, the total effects of strategy use on reading test performance are still limited. This suggests that the contributions of students' strategy deployment to their reading test performance are finally small.

As can be seen in Table 4.4, the *monitoring, directing attention and managing the test* (MDAMT) process yields the most *positive*, total effects on multiple-choice reading comprehension test performance (MC RCTP), whereas the *monitoring and utilizing test questions* (MUTQ) process has a *negative*, total effect. The results suggest that in this reading test-taking context, these EFL students are capable of appropriately deploying monitoring, repeating, retrieving-linking and managing-the-test strategies covered by the MDAMT process, so that the MDAMT process makes the most contributions to their reading test performance among all strategy use processes. However, they are unable to aptly employ monitoring and taking-advantage-of-test-questions strategies contained by the MUTQ process, given that the MUTQ process is slightly detrimental to their reading test performance.

As expected, compared with reading and test-taking strategy use, English language knowledge overall yields more effects on multiple-choice reading comprehension test performance (see Table 4.4). This demonstrates that students' English language knowledge is more profitable to how well they perform the reading test than is their strategy use.

It is worth noting that during the model production process, the modification indices indicated that explicit questions (ExQ) had an impact on the *monitoring the reading process with positive results* (MRPPR) strategy subgroup. Given the parsimony and interpretability of the overall model, I did not take into account and accept this causal effect at that moment. However, a later reflection shows that such information seems reasonable. Explicit questions in the multiple-choice reading test provide information related to reading passages of which students make sense. Such information probably facilitates students' deploying partial monitoring strategies subsumed by the MRPPR strategy subgroup. It then follows that ExQ exerts an effect on the MRPPR strategy subgroup. If this is the case, the current study provides empirical evidence for Bachman and Palmer's (1996) model of language ability in language use and language test performance regarding the interactive relationship between test-takers' characteristics and attributes of test tasks. To illustrate, this accepted model manifests that English language knowledge and strategy use yield an effect on multiple-choice reading comprehension test performance, which can be regarded as an example that test-taker characteristics affect test tasks. If the effect of ExQ on the MRPPR strategy subgroup were adopted, this could be thought of as an illustration that test tasks influence test-taker characteristics. It follows that test-taker characteristics interact with test tasks, as shown in Bachman and Palmer's (1996) model of language ability in language use and language test performance. A further study with different reading tasks is needed to give empirical evidence for this interactive relation.

In summary, within this multiple-choice reading test context, students' English language knowledge exercises more influences on their reading test performance than does strategy use. The effect that their strategy use on reading test performance is limited and not as strong as expected. In addition, students' strategy use has trivial or weak, indirect, positive impacts on their reading test performance by means of either English language knowledge or English language knowledge and other strategy use. On the other hand, their English language knowledge yields a trivial, indirect, positive effect on reading test performance through strategy use.

4.4 Conclusion

This chapter addresses the results regarding the relationship among students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance. Within the accepted SEM model, neither students' English language knowledge nor their reading and test-taking strategy use is a single-dimensional construct. Both students' English language knowledge and strategy deployment have an effect on their reading test performance with divergent effect strengths and in multi-directional manners.

Within English language knowledge, students' grammatical knowledge exerts more effects on their reading test performance than does lexical knowledge. Their grammatical knowledge also indirectly affects their reading test performance by means of lexical knowledge or strategy deployment.

Students' strategy utilization does not always yield a positive effect on reading test performance. Their *monitoring and utilizing test questions* process shows a trivial, negative effect on their reading test performance. Among all strategy use processes, students' *monitoring, directing attention and managing the test* process has the most impacts on their reading test performance. Further, some of students' strategy use impacts upon their reading test performance through other strategy deployment, English language knowledge, or English language knowledge and other strategy employment.

Finally, in comparison with strategy use, students' English language knowledge shows more effects on their reading test performance. Their English language knowledge and strategy use interact with each other in the reading test-taking context.

The following chapter will be concerned with the multiple group analyses. Grounded on the accepted full latent variable model for the entire group, I will hypothesize and test the separate models regarding the relation among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the high English ability group and the low English ability group. In addition, I will perform the simultaneous group analysis for both group models.

CHAPTER FIVE

MUTIPLE GROUP ANALYSIS: RESULTS

5.1 Introduction

This chapter further explores to what extent the model concerning the relation amongst English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance varies across English ability levels. The focus is on comparisons of and invariance tests for the two separate full latent variable models that profile the aforementioned relationship for two groups with different English ability. According to the result of their self-rated English ability, participants were divided into two groups: the high English ability (HEA) group and the low English ability (LEA) group. Then, the separate full latent variable models regarding the relation among English language knowledge, strategy use, and reading test performance for the HEA and the LEA groups were hypothesized and tested. For cross-group invariance tests, these two group models were also estimated simultaneously with equality constraints imposed on parameters of interest to test whether these parameters were equivalent across English ability levels.

This chapter is structured as follows. To begin with, I explain how participants were divided into two groups: the HEA group and the LEA group. Then, I describe how the relationship among English language knowledge, strategy use and reading test performance were constructed for these two groups. The individual models for the two groups are compared and discussed briefly. Finally, I discuss the cross-group invariance tests and the results of the tests.

5.2 The high English ability group and the low English ability group

With regard to the criterion for separating participants into groups, it first came to my mind that participants were divided into two groups, based on the scores of the English language knowledge test and the reading comprehension test. However, a later reflection showed that it appeared inadequate. The reason is as follows. If the English language knowledge test scores or the reading comprehension test scores are used for group division, the discrepancies in reading test performance or in English language knowledge across groups with different English ability will be observed expectedly, given that the scores of the English language knowledge test and the reading comprehension test are also submitted to the structural equation modeling analysis. Then, such results are unreliable to a certain degree, seeing that they are attributed to a researcher's obvious manipulation.

It follows that it is more appropriate to administer alternative tests to gauge participants' English ability and adopt the scores of the tests to categorize them into groups. However, given the large sample size (N = 1064), the limited data collection time, and the unavailability of appropriate tests, it is not feasible. Therefore, adopting other methods, rather than administering alternative tests to participants, is preferred. The self-rating method has been adopted in previous L2 studies to assess participants' English ability or reading ability (e.g., Oxford & Nyiko, 1989; Sheorey, 1999; Sheorey & Edit, 2004; Sheorey & Mokhtari, 2001) and the result has been used to divide participants into groups in some studies (e.g., Sheorey & Edit, 2004). Accordingly, in the current study, I utilized the score of self-rated English ability as a criterion for group division.

Participants who self-rated their English ability 13 or above (out of 20) were classified into the high English ability (HEA) group, whilst those who self-rated their English ability 12 or below were categorized into the low English ability (LEA) group. The HEA group consisted of 322 participants, while the LEA group constituted 512 participants.

5.3 A t-test for the HEA group and the LEA group

As suggested by Kunnan (1998), prior to postulating the separate full latent variable models for different groups, it is necessary to assess whether there are significant differences across groups in the variables of interest from a statistical perspective. If not, then there is no need to construct individual full latent variable models for divergent groups. As a consequence, I conducted an independent samples t-test to examine whether significant differences existed in English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance between the HEA and the LEA groups, previous to constructing two full latent variable models for these two groups.

The result of the t-test showed that despite the limited mean differences, significant differences ($p < .050$) were present between the HEA and the LEA groups in (a) all language-knowledge-test-item subgroups (i.e., LEX1, LEX2, GRAM1 and GRAM2); (b) both language-knowledge-type subgroups (i.e., lexical knowledge and grammatical knowledge); (c) the entire English language knowledge test; (d) all strategy subgroups (e.g., the *repeating* strategy subgroup); (e) all strategy use processes (e.g., the *evaluating and marking* process); (f) overall strategy use; (g) both reading-test-item subgroups (i.e., explicit questions and inferential questions); and (h) the entire reading test (for the results in detail, see Appendix 12). The results provide support for the subsequent analysis – two individual English-ability-group models regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance are constructed and compared to pinpoint cross-group commonalities and variations.

5.4 Constructing the separate full latent variable models regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the HEA group and the LEA group

The same procedures carried out in the entire group analysis were applied to the separate group analyses. Based on the measurement models of English language knowledge and reading and test-taking strategy use for the entire group, I began to posit and test these two measurement models for the HEA and the LEA groups with the conduction of confirmatory factor analyses (CFAs). The result showed that the measurement models of English language knowledge for these two groups were equivalent. For either case, the four test-item subgroups (observed variables) well explained lexical knowledge and grammatical knowledge (latent variables), with factor loadings ranging from .740 to .850 for the HEA group and from .721 to .831 for the LEA group (for details, see Appendices 13 and 16). As for the measurement model of reading and test-taking strategy use, there were some differences across the groups, which were discussed in Section 5.4.1.2. Except for cross-loadings, the twelve strategy subgroups (observed variables) adequately accounted for the four strategy use processes (latent variables), with factor loadings varying from .513 to .822 for the HEA group and from .569 to .956 for the LEA group (for details, see Appendices 14 and 17). I did not submit the reading comprehension test performance to CFAs, given that the number of observed variables is limited and SEM cannot be performed for CFAs. Then, I modeled and tested the full latent variable models concerning the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance respectively for the HEA and the LEA groups grounded on the entire group model to address the following research questions.

2. Is there a difference in the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance across English ability levels?
   2.1 Is there a difference in students' English language knowledge and reading and test-taking strategy use contributing to their multiple-choice reading comprehension test performance across English ability levels? Do the relative contributions of students' English language knowledge and reading and test-taking strategy use to their multiple-choice reading comprehension test performance differ across English ability levels?
   2.2 Is there a language threshold for students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance across English ability levels?
   2.3 Is there a difference in the relationship between students' English language knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test across English ability levels?

5.4.1 The results for the separate full latent variable models for the HEA group and the LEA group

After forty-six runs for model testing and model repsecification, two SEM models with fair goodness-of-fit and interpretability were produced for the HEA and the LEA groups (see Appendices 15 and 18 for the accepted models). In this section, the focus is on examining the model fit statistics of both group models to justify the appropriateness for accepting the two models. Table 5.1 depicts the model fit indices for the two group models.

Table 5.1 The model fit indices for the full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

Model fit indices	Levels of acceptable fit	Evaluation results	
		The HEA Group	The LEA Group
χ^2	Nonsignificant with the p-value > .050	Good (109.505 with p = .576 > .050)	Good (110.855 with p = .459 > .050)
GFI	> .900	Very good (GFI = .962)	Very good (GFI = .975)
AGFI	> .900	Good (AGFI = .943)	Very good (AGFI = .961)
CFI	> .950	Very good (CFI = 1.000)	Very good (CFI = 1.000)
TLI	> .950	Very good (TLI = 1.002)	Very good (TLI = 1.000)
RMSEA	< .060	Very good (RMSEA = .000)	Very good (RMSEA = .004)

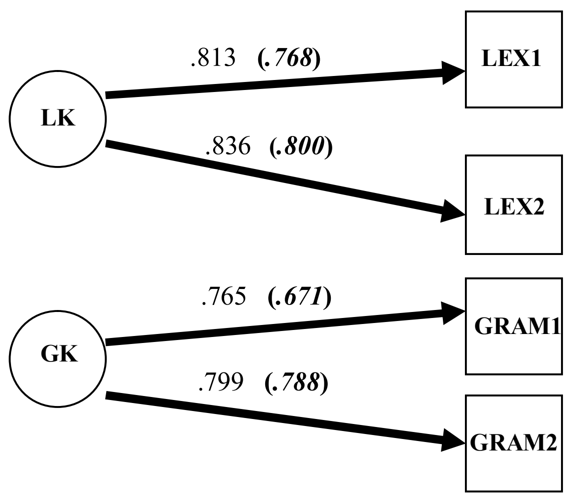
Note. n=300 for the HEA group; n=476 for the LEA group. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index, CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

As seen in Table 5.1, the chi-square statistics of 109.505 for the HEA group and 110.855 for the LEA group were not statistically significant at the .050 level: their p-values were .576 for the HEA group and .459 for the LEA group. The values of the GFI, the AGFI, the CFI and the TLI respectively corresponded to .962, .943, 1.000 and 1.002 for the HEA group and .975, .961, 1.000 and 1.000 for the LEA group. All of these indices in both groups exceeded the cut-off value. Similarly, the RMSEA results of .000 for the HEA group and .004 for the LEA group were below the .060 threshold. Based on the model fit indices shown here, the models for both groups described the sample data satisfactorily and it was adequate to accept them. All the effect paths and correlations listed in both models were statistically significant at the .050 level ($p < .050$). In addition, I performed the bootstrap analysis to inspect whether indirect effects manifested in both models were statistically significant. The result of the bootstrap analysis indicated that all the indirect effects reached statistical significance ($p < .050$) for either case, except for the indirect effects of the *monitoring and utilizing test questions* (MUTQ) process on explicit questions (ExQ) and inferential questions (InQ) in the LEA group. In the following sections, I will first examine the measurement models of English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the HEA and the LEA groups. Then, I shift to the structural model regarding the relation amongst English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance.

5.4.1.1 The measurement model of English language knowledge for the HEA group and the LEA group: The separate group analysis

In this section, I first outline what the measurement model of English language knowledge comprises for the HEA and the LEA groups. Then, I examine factor loadings shown in this measurement model for both groups.

Within the measurement model of English language knowledge (ELK), the equivalent component structures are observed for the HEA and the LEA groups. More specifically, the measurement model of ELK in either level is symbolized by two latent variables: lexical knowledge (LK) and grammatical knowledge (GK). Both LK and GK are respectively represented by two markers: the LEX1 and the LEX2 test-item subgroups for LK; the GRAM1 and the GRAM2 test-item subgroups for GK.



LK=Lexical knowledge; GK=Grammatical knowledge. LEX1 consists of ten test items of the vocabulary subtest; LEX2 eleven test items of the vocabulary subtest; GRAM1 eight test items of the grammar subtest; GRAM2 eight test items of the grammar subtest; Factor loadings for the HEA group are not in the parenthesis, while those for the LEA group are in the parenthesis.

→ =Observed variables load on latent variables.

Figure 5.1 The measurement model of English language knowledge for the HEA group and the LEA group: The separate group analysis

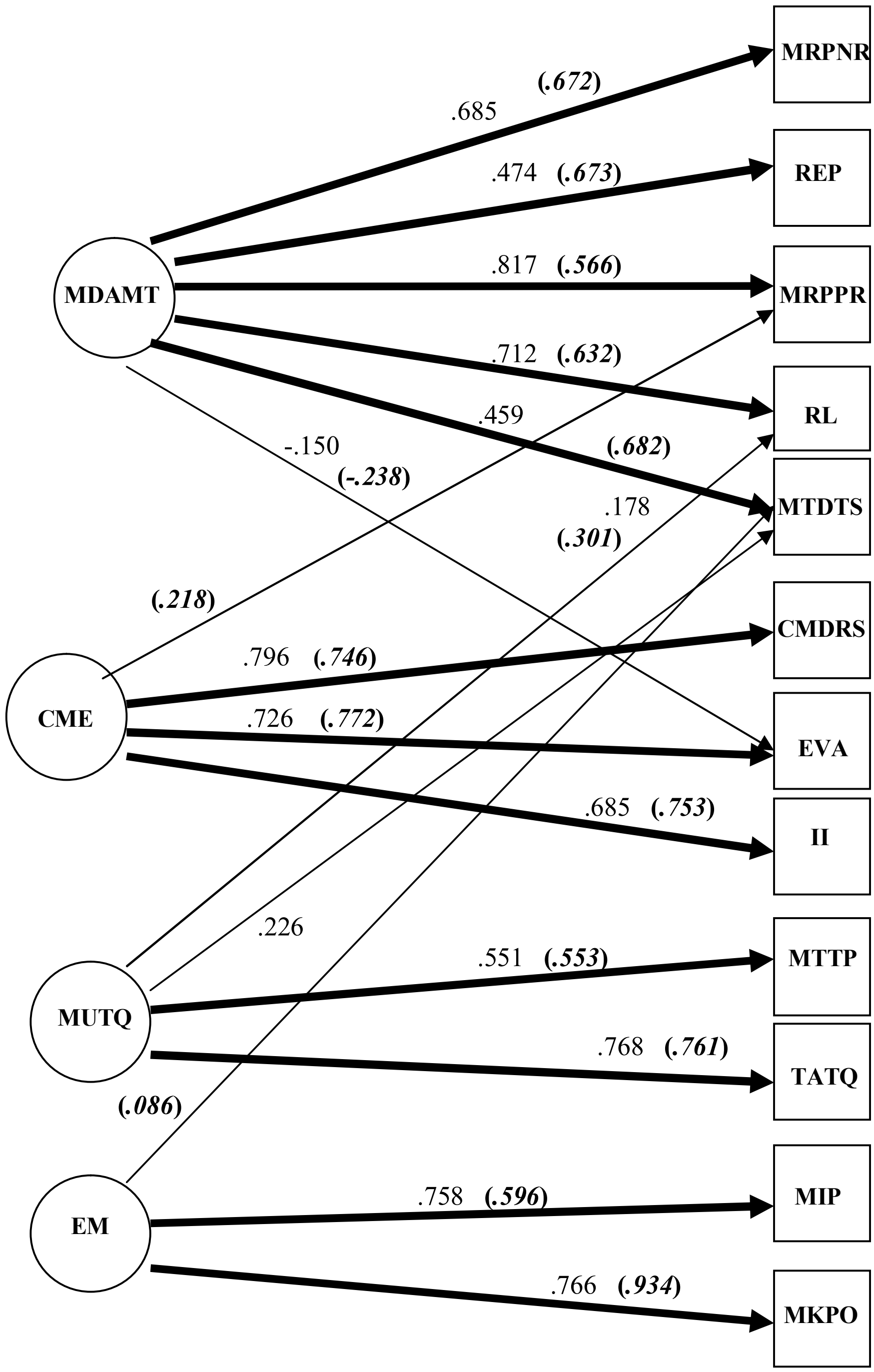
As presented in Figure 5.1, in the two groups, the LEX1, LEX2, GRAM1, and GRAM2 test-item subgroups respectively explain lexical knowledge (LK) and grammatical knowledge (GK) well, with loadings ranging from .765 to .836 for the HEA group and .671 to .800 for the LEA group. Surprisingly, the factor loadings that the LEX1 and LEX2 test-item subgroups produce on LK, and the GRAM1 and GRAM2 test-item subgroups on GK are similar between the HEA and the LEA groups. The HEA group possess better English language knowledge than the LEA group. The LEX1 and LEX2 test-item subgroups should measure LK better in the HEA group than in the LEA group and so the GRAM1 and GRAM2 test-item subgroups in GK. Then, the LEX1 and LEX2 test-item subgroups should yield more factor loadings on LK and the GRAM1 and GRAM2 test-item subgroups on GK in the HEA group than in the LEA group. The similar factor loadings between the two groups within this measurement model can be partially attributed to the limited difference in the score of the English language knowledge test across the groups – the mean of the English language knowledge test for the HEA group was 27.000, while that for the LEA group was 23.272.

In summary, for the HEA and the LEA groups, within the measurement model of English language knowledge, the LEX1 and LEX2 test-item subgroups well gauge lexical knowledge. Similarly, the GRAM1 and GRAM2 test-item subgroups well assess grammatical knowledge.

5.4.1.2 The measurement model of reading and test-taking strategy use for the HEA group and the LEA group: The separate group analysis

In this section, I begin to briefly describe what the measurement model of reading and test-taking strategy use encompasses for both groups. Next, I inspect factor loadings and cross-loadings manifested in this measurement model for either case.

Similar to that in the entire group, four components (strategy use processes) represent reading and test-taking strategy use in either group. These components consist of the *monitoring, directing attention and managing the test* (MDAMT), the *constructing the meaning and evaluating* (CME), the *monitoring and utilizing test questions* (MUTQ) and the *evaluating and marking* (EM) processes.



MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ= Monitoring and utilizing test questions; EM=Evaluating and marking; MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA= Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP=Marking incomprehensible parts; MKPO=Marking key points or options. Factor loadings for the HEA group are not in the parenthesis, while those for the LEA group are in the parenthesis. ➡ =Observed variables load on latent variables. ⟶ =Observed variables cross-load on latent variables.

Figure 5.2 The measurement model of reading and test-taking strategy use for the HEA group and the LEA group: The separate group analysis

Like the overall group, the *monitoring, directing attention and managing the test* (MDAMT) process, for both groups, is appropriately explained by the *monitoring the reading process with negative results* (MRPNP), the *repeating* (REP), the *monitoring the reading process with positive results* (MRPPR), the *retrieving-linking* (RL) and the *managing the test with the deployment of test-taking strategies* (MTDTS) strategy subgroups. The loadings range from .459 to .817 for the HEA group and .566 to .682 for the LEA group. A comparison between the HEA group and the LEA group shows that the HEA group differs from the LEA group in three strategy subgroups. The *repeating* (REP) and the *managing the test with the deployment of test-taking strategies* (MTDTS) strategy subgroups produce more loadings on the MDAMT process in the LEA group than in the HEA group (.673 vs. .474 for REP; .682 vs. .459 for MTDTS). In contrast, the *monitoring the reading process with positive results* (MRPPR) strategy subgroup yields more loadings on the MDAMT process in the HEA group model (.817), compared with that in the LEA group model (.566). These findings suggest discrepancies between these two groups in the deployment of repeating strategies, partial monitoring strategies, and managing-the-test strategies within the MDAMT process. These cross-group differences were further tested in the simultaneous group analysis.

Within the *constructing the meaning and evaluating* (CME) process, in either group, the CME process is fairly accounted for by the *constructing the meaning with the deployment of reading strategies* (CMDRS), the *evaluating* (EVA) and the *interacting with the input* (II) strategy subgroups. The factor loadings vary from .685 to .796 for the HEA group and .746 to .772 for the LEA group. When making a between-group comparison, I found that these factor loadings were similar across the groups, which ran counter to my expectation. The factor loadings in the HEA group should be stronger or less strong than those in the LEA group given a variation in reading strategy use across English ability levels, as previous reading strategy studies suggest (e.g., Davis & Bistodeau, 1993; Oxford, *et al*., 2004; Upton & Lee-Thompson, 2001). The cross-group similarity in the factor loadings for the CME process indicates that perhaps the HEA group rely upon constructing-the-meaning and evaluating strategies encompassed by the CME process to a similar extent as the LEA group in this reading test-taking context.

Similarly, the *monitoring and utilizing test questions* (MUTQ) process is properly gauged by its indicators, the *monitoring the test-taking process* (MTTP) and the *taking advantage of test questions* (TATQ) strategy subgroups. The loadings are .551 and .768 for the HEA group and .553 and .761 for the LEA group.

With regard to the *evaluating and marking* (EM) process, the *marking incomprehensible parts* (MIP) and the *marking key points or options* (MKPO) strategy subgroups, in either case, serve as fair indicators, with loadings of .758 and .766 for the HEA group and .596 and .934. When a between-group comparison was made, I found something interesting. The loading that the *marking incomprehensible parts* (MIP) strategy subgroup produces on the EM process in the HEA group is larger than that in the LEA group (.758 vs. .596). Such a finding suggests that marking-incomprehensible-parts strategies seem to promote the EM process more in the HEA group than the case in the LEA group. The HEA group appear to make more effort to make sense of the input than the LEA group. On the other hand, the loading that the *marking key points or options* (MKPO) strategy subgroup yields on the EM process in the HEA group is smaller than that in the LEA group (.766 vs. .934). This implies that marking-key-points-or-options strategies less contributes to the EM process in the HEA group than the case in the LEA group. These cross-group discrepancies were further tested in the simultaneous group analysis.

While cross-component loadings are observed in both group models, there are variations in the number of and the types of cross-component loadings. More specifically, the HEA group model captures three cross-loadings, whereas the LEA group model four. The HEA group do not share all cross-component loadings with the LEA group. As regards cross-component loadings shared by the two levels, the *interacting with the input* (II) strategy subgroup cross-loads on the *monitoring, directing attention and managing the test* (MDAMT) process with a negative loading (-.150 for the HEA group and -.238 for the LEA group). In addition, in either level, the *retrieving-linking* (RL) strategy subgroup yields a cross-loading on the *monitoring and utilizing test questions* (MUTQ) process with a weak or moderate positive loading (.178 for the HEA group and .301 for the LEA group). Notice that these cross-loadings in the LEA group are larger than those in the HEA group. This reveals the presence of cross-group differences, even though both groups draw upon the identical type of strategy within the same strategy use process.

Aside from the same cross-component loadings between the groups, there are different cross-component loadings across these two groups. For example, in the HEA group model, the *managing the test with the deployment of test-taking strategies* (MTDTS) strategy subgroup generates a loading on the *monitoring, directing attention and managing the test* (MDAMT) process (.459) and the *monitoring and utilizing test questions* (MUTQ) process (.226). The result suggests that the HEA group deploy managing-the-test strategies to aid them in monitoring the reading process, administering the overall test and taking advantage of test questions in this test-taking setting.

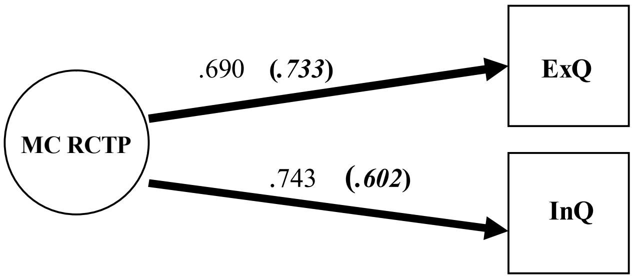
Unlike that in the HEA group, within the LEA group, the *monitoring the reading process with positive results* (MRPPR) strategy subgroup yields a loading on the *monitoring, directing attention and managing the test* (MDAMT) process (.566) and the *constructing the meaning and evaluating* (CME) process (.218). The result implies that the LEA group employ partial monitoring strategies to oversee the reading process, manage the test, as well as grasp the text. On the other hand, the HEA group deploy these strategies uniquely to supervise the reading process and administer the test.

In summary, for the HEA and the LEA groups, strategy subgroups overall appropriately explain strategy use processes in the measurement model of reading and test-taking strategy use, despite the presence of cross-loadings. Although a similar component structure is shared between these two groups, how the underlying observed variables (i.e., strategy subgroups) perform is not fully the same across the groups. In other words, there are certain discrepancies in strategy employment across English ability levels, as suggested in several strategy-related studies (e.g., Nikolov, 2006; Oxford *et al*., 2004; Purpura, 1998b; 1999; Upton & Lee-Thompson, 2001).

5.4.1.3 The measurement model of multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

In this section, at first I briefly depict what the measurement model of multiple-choice reading comprehension test performance subsumes for the HEA and the LEA groups. Then, I examine factor loadings observed in this measurement model for either case.

With regard to the measurement model of multiple-choice reading comprehension test performance (MC RCTP), the HEA group share with the LEA group a commonality in the component structure of MC RCTP. Like the entire group, for these two groups, MC RCTP (a latent variable) is characterized by two indicators: explicit questions (ExQ) and inferential questions (InQ).



MC RCTP=Multiple-choice reading comprehension test performance; ExQ=Explicit questions; InQ=Inferential questions. Factor loadings for the HEA group are not in the parenthesis, while those for the LEA group are in the parenthesis. → =Observed variables load on latent variables.

Figure 5.3 The measurement model of multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

As illustrated in Figure 5.3, explicit questions (ExQ) and inferential questions (InQ), either in the HEA group or in the LEA group models, moderately account for multiple-choice reading comprehension test performance (MC RCTP), with loadings of .690 and .743 for the HEA group, while .733 and .602 for the LEA group. When further inspecting these loadings between the groups, I found a piece of intriguing information. The loading that inferential questions (InQ) generate on multiple-choice reading comprehension test performance (MC RCTP) in the HEA group is larger than that in the LEA group (.743 vs. .602). This indicates that InQ functions as a better indicator for MC RCTP in the HEA group than in the LEA group. The finding is reasonable as one can presume that inferential questions entail students' comprehending reading passages thoroughly and the HEA group, compared with the LEA group, are more able to make comprehensive sense of reading passages. This cross-group difference was further tested in the simultaneous group analysis.

To sum up, for the HEA and the LEA groups, both explicit and inferential questions serve well as indicators for multiple-choice reading comprehension test performance. In addition, while both groups share the equivalent component structure of multiple-choice reading comprehension test performance with each other, how the underlying observed variables (i.e., explicit questions and inferential questions) perform is not be completely identical across the groups.

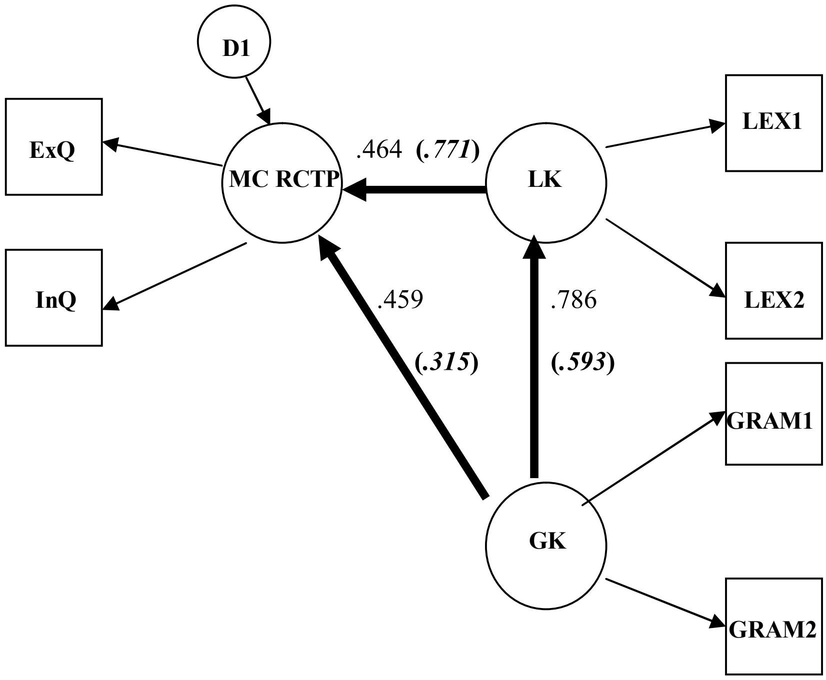
5.4.1.4 The relationship between English language knowledge and multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

In this section, I first examine an effect that lexical knowledge exerts on multiple-choice reading comprehension test performance for both groups and then turn to an impact that grammatical knowledge has on multiple-choice reading comprehension test performance.

As can be seen in Figure 5.4, the equivalent structural relationship between English language knowledge (ELK) and multiple-choice reading comprehension test performance (MC RCTP) is shared across the two levels but with different effect strengths. Within the HEA group model, lexical knowledge (LK) displays a *moderate*, direct, positive effect on MC RCTP, with a value of .464. By contrast, in the LEA group model, LK yields a *strong*, direct, positive effect on MC RCTP, with a value of .771. Such a difference suggests that the LEA group, in comparison with the HEA group, appear to rest more on their lexical knowledge to deal with the reading comprehension test. This cross-group difference was further tested in the simultaneous group analysis.

Distinct from lexical knowledge, in either group, grammatical knowledge (GK) yields a moderate, direct, positive effect on multiple-choice reading comprehension test performance (MC RCTP), with values of .459 for the HEA group and .315 for the LEA group. Notice that the direct effect that MC RCTP receives from GK in the HEA group is stronger than that in the LEA group, suggesting a possible variation in the extent to which GK directly impacts upon MC RCTP between the groups.

When closely inspecting direct effects of lexical knowledge (LK) and grammatical knowledge (GK) on multiple-choice reading comprehension test performance (MC RCTP), I found an interesting cross-group discrepancy. In the HEA group, LK and GK *directly* impact upon MC RCTP almost equally (.464 vs. .459). The HEA group, as sitting this reading comprehension test, seem to directly depend on lexical knowledge as heavily as grammatical knowledge. The situation differs in the LEA group. LK *directly* influences MC RCTP even more than GK (.771 vs. 315). Unlike the HEA group, the LEA group appear to directly rely upon lexical knowledge more than grammatical knowledge to tackle the reading comprehension test.



MC RCTP=Multiple-choice reading comprehension test performance; LK=Lexical knowledge; GK=Grammatical knowledge. Effects for the HEA group are not in the parenthesis, while those for the LEA group are in the parenthesis. ➡ =A latent variable has an effect on another latent variable.

Figure 5.4 The relationship between English language knowledge and multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

Table 5.2 Effects of English language knowledge on multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

	Effects	
	The HEA Group	The LEA Group
LK→MC RCTP	.464	.771
GK→MC RCTP[a]	.459	.315
GK→LK→ MC RCTP[b]	.365	.457

Note. MC RCTP=Multiple-choice reading comprehension test performance; LK=Lexical knowledge; GK=Grammatical knowledge. a=A direct effect of GK on MC RCTP; b=An indirect effect of GK on MC RCTP.

In addition to direct effects, both group models capture the indirect effect of grammatical knowledge (GK) on multiple-choice reading comprehension test performance (MC RCTP) by means of lexical knowledge (LK), with values of .365 for the HEA group and .457 for the LEA group (see Table 5.2). The result suggests the presence of the indirect contribution that the HEA and the LEA groups' grammatical knowledge make to their reading test performance. It is somewhat curious that the indirect effect within the HEA group is slightly smaller than that in the LEA group. The HEA group are equipped with more grammatical knowledge than the LEA group. They should be more capable than the LEA group of accessing grammatical knowledge to activate lexical knowledge to tackle the reading test. Then, the effect of grammatical knowledge on reading test performance through lexical knowledge should be larger in the HEA group than that in the LEA group. However, this expectation is not manifested in the current data set. The reason may rest on the fact that the HEA group access grammatical knowledge to activate their lexical knowledge to deal with the reading test in a more automatic manner than the LEA group.

To summarize, the HEA and the LEA groups' lexical knowledge as well as grammatical knowledge shows a positive effect on their multiple-choice reading test performance. However, there are differences across the English ability levels. The HEA group's grammatical knowledge and lexical knowledge yield similar direct effects on their reading test performance. In contrast, the LEA group's lexical knowledge exercises more direct influences on their reading test performance than grammatical knowledge. Additionally, grammatical knowledge in the HEA group has more direct impacts on reading test performance than that in the LEA group. Finally, lexical knowledge in the LEA group exerts more direct effects on reading test performance, compared with the case in the HEA group.

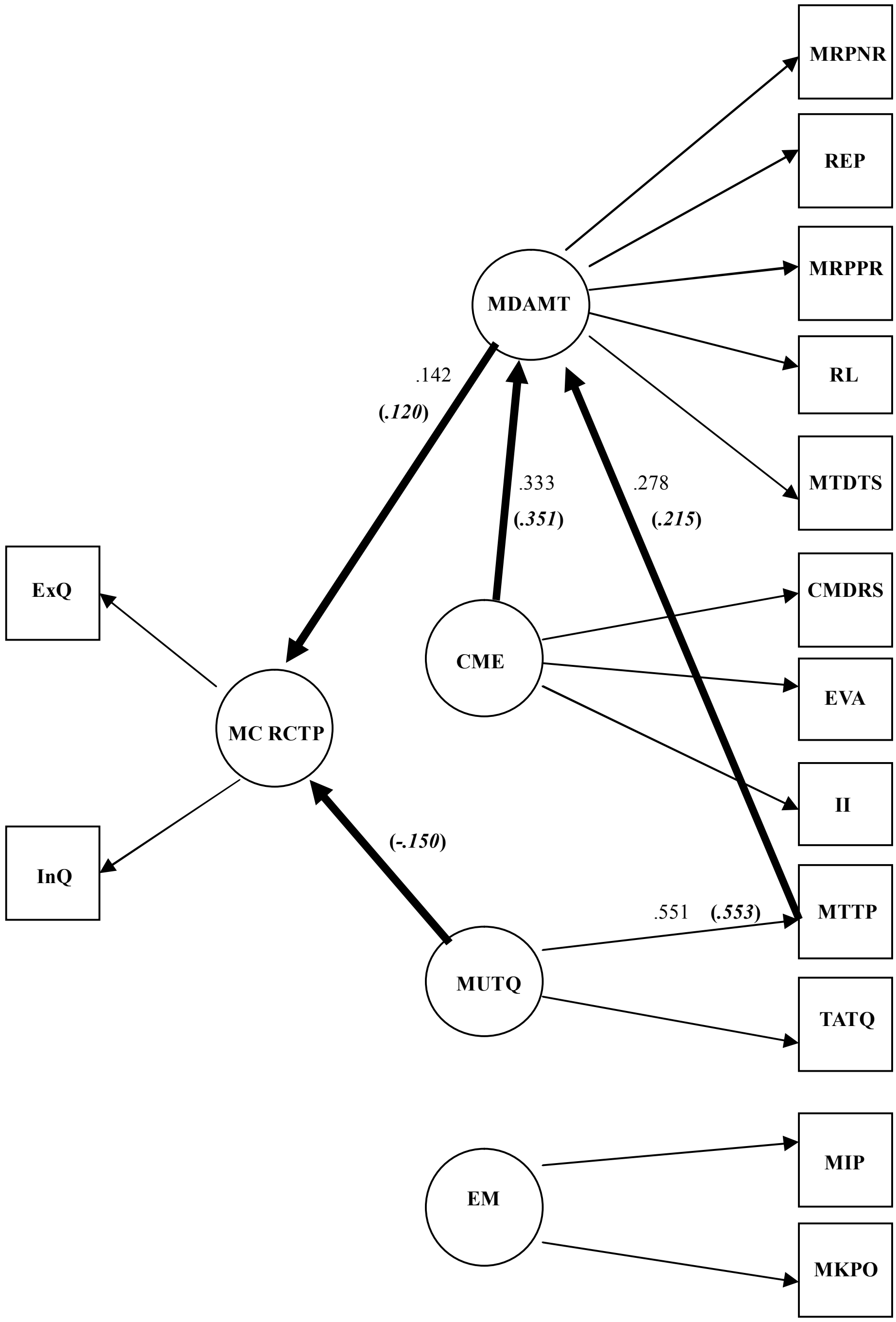
5.4.1.5 The relationship between reading and test-taking strategy use and multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

In this section, I examine an effect that reading and test-taking strategy use (i.e., strategy use processes and strategy subgroups) has on multiple-choice reading comprehension test performance in the HEA and the LEA group models.

With regard to the relationship between strategy use and multiple-choice reading comprehension test performance (MC RCTP), similar paths are manifested for both groups. Like the case in the entire group model, not all strategy use has an impact on multiple-choice reading comprehension test performance (MC RCTP). Both group models indicate that the *evaluating and marking* (EM) process yields no significant, direct or indirect effect on MC RCTP. The result appears to suggest that both groups' employment of marking strategies exercises no noticeable influence on how well these two groups perform the reading test.

Distinct from the *evaluating and marking* (EM) process, within either group, the *monitoring, directing attention and managing the test* (MDAMT) process has a weak, direct, positive impact on multiple-choice reading comprehension test performance (MC RCTP), with values of .142 for the HEA group and .120 for the LEA group. Notice that the direct effect that the MDAMT process shows on MC RCTP is similar across English ability levels. The result is somewhat curious. I originally expected that more direct effects were observed in the HEA group, given that the HEA group's strategy use, generally, is more effective than the LEA group's. Such an expectation, however, is not manifested within the current data set. Maybe some of the HEA group's employment of monitoring, repeating, retrieving-linking and managing-the-test strategies covered by the MDAMT process is in an automatic way, so that the less strong effect of the MDAMT process on MC RCTP is captured in their group model. Then, a similarity is present in the direct effect that the MDAMT process shows on MC RCTP across the HEA and the LEA groups.

Similar to that in the entire group model, the *constructing the meaning and evaluating* (CME) process in both group models yields a trivial, *indirect*, positive effect on multiple-choice reading comprehension test performance (MC RCTP), by means of the *monitoring, directing attention and managing the test* process, with values of .047 for the HEA group and .042 for the LEA group (see Table 5.3).



MC RCTP=Multiple-choice reading comprehension test performance; MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; EM=Evaluating and marking; MTTP=Monitoring the test-taking process. Effects for the HEA group are not in the parenthesis, while those for the LEA group are in the parenthesis. ➡ =A latent variable or an observed variable has an effect on another latent variable.

Figure 5.5 The relationship between reading and test-taking strategy use and multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

Table 5.3 Effects of reading and test-taking strategy use on multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

	Effects	
	The HEA Group	The LEA Group
MDAMT→MC RCTP	.142	.120
CME→MDAMT→MC RCTP	.047	.042
MTTP→MDAMT→MC RCTP	.040	.026
MUTQ→MC RCTP[a]	No	-.150
MUTQ→MTTP→MC RCTP[b]	.022	.014

Note. MC RCTP=Multiple-choice reading comprehension test performance; MDAMT= Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; EM=Evaluating and marking; MTTP=Monitoring the test-taking process. a=A direct effect of MUTQ on MC RCTP; b=An indirect effect of MUTQ on MC RCTP.

On further examining the relation between strategy use and reading test performance, I unexpectedly found that a strategy subgroup (an observed variable) displayed an effect on multiple-choice reading comprehension test performance (MC RCTP). To explain, the *monitoring the test-taking process* (MTTP) strategy subgroup has a trivial, indirect, positive impact on MC RCTP by means of the *monitoring, directing attention and managing the test* (MDAMT) process, with values of .040 for the HEA group and .026 for the LEA group (see Table 5.3). The result suggests that both groups' employment of the strategies related to supervising the overall test-taking process is able to be trivially, indirectly beneficial to their performance on the reading test, through their deployment of monitoring, repeating, retrieving-linking, and managing-the-test strategies subsumed by the MDAMT process. Moreover, the result here, coupled with that mentioned above, implies that sometimes the HEA and the LEA groups' strategy use necessitates other strategy deployment with a view to enhancing their task performance.

A discrepancy across these two groups is reflected in the effect of the *monitoring and utilizing test questions* (MUTQ) process on multiple-choice reading comprehension test performance (MC RCTP). Within the HEA group, the MUTQ process does not exert any direct influence on MC RCTP. Neither does it have any indirect impact on MC RCTP by means of other strategy use processes. Rather, through the *monitoring the test-taking process* strategy subgroup, the MUTQ process yields a trivial, *indirect*, *positive* effect on

MC RCTP, with a value of .022 (see Table 5.3). The result suggests that the HEA group's deploying monitoring and taking-advantage-of-test-questions strategies slightly contributes to their reading comprehension test performance in an indirect fashion.

The LEA group shows a different story. The *monitoring and utilizing test questions* (MUTQ) process has a weak, *direct*, *negative* influence on multiple-choice reading comprehension test performance (MC RCTP), with a value of -.150. The result concurs with that in the overall group analysis. The LEA group's employment of monitoring and utilizing-test-questions strategies is weakly, directly detrimental to their reading test performance. This indicates that the LEA group's strategy deployment, unlike the HEA group's, probably has not developed to a stage at which their strategy use can always promote their reading test performance in this test-taking context.

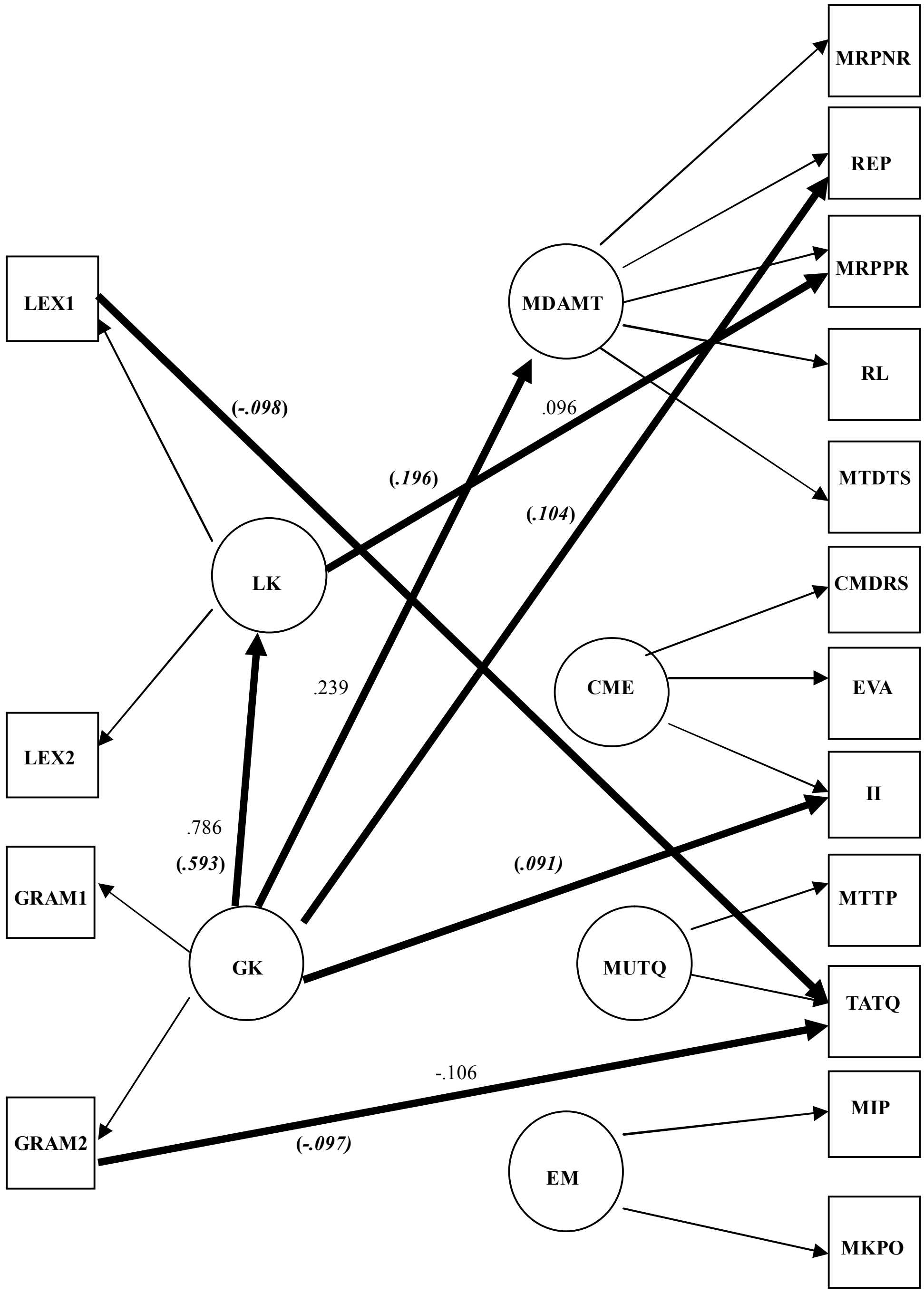
Finally, a between-group comparison indicates that direct effects or indirect effects of strategy use on reading test performance in the HEA group model are overall slightly stronger than those in the LEA group model. Furthermore, distinct from that in the LEA group model, no direct, negative effect is present in the HEA group model. These findings suggest that the HEA group deploy strategies in a more effective and appropriate manner than the LEA group in this reading test situation, given that the HEA group's strategy employment, on the whole, makes slightly more contributions to reading test performance than the LEA group's strategy use.

In summary, the HEA and the LEA groups' strategy deployment, although not all, exercises an influence on their multiple-choice reading comprehension test performance to a certain degree. For either group, the *monitoring, directing attention and managing the test* process, the *constructing the meaning and evaluating* process, and the *monitoring the test-taking process* strategy subgroup display trivial or weak, positive effects on reading test performance. Further, the *monitoring and utilizing test questions* process has a weak, direct, *negative* impact on multiple-choice reading comprehension test performance in the LEA group. However, such an effect is absent in the HEA group. Clearly, strategy deployment varies to some extent across these two groups in the L2 reading test-taking context.

5.4.1.6 The relationship between English language knowledge and reading and test-taking strategy use for the HEA group and the LEA group: The separate group analysis

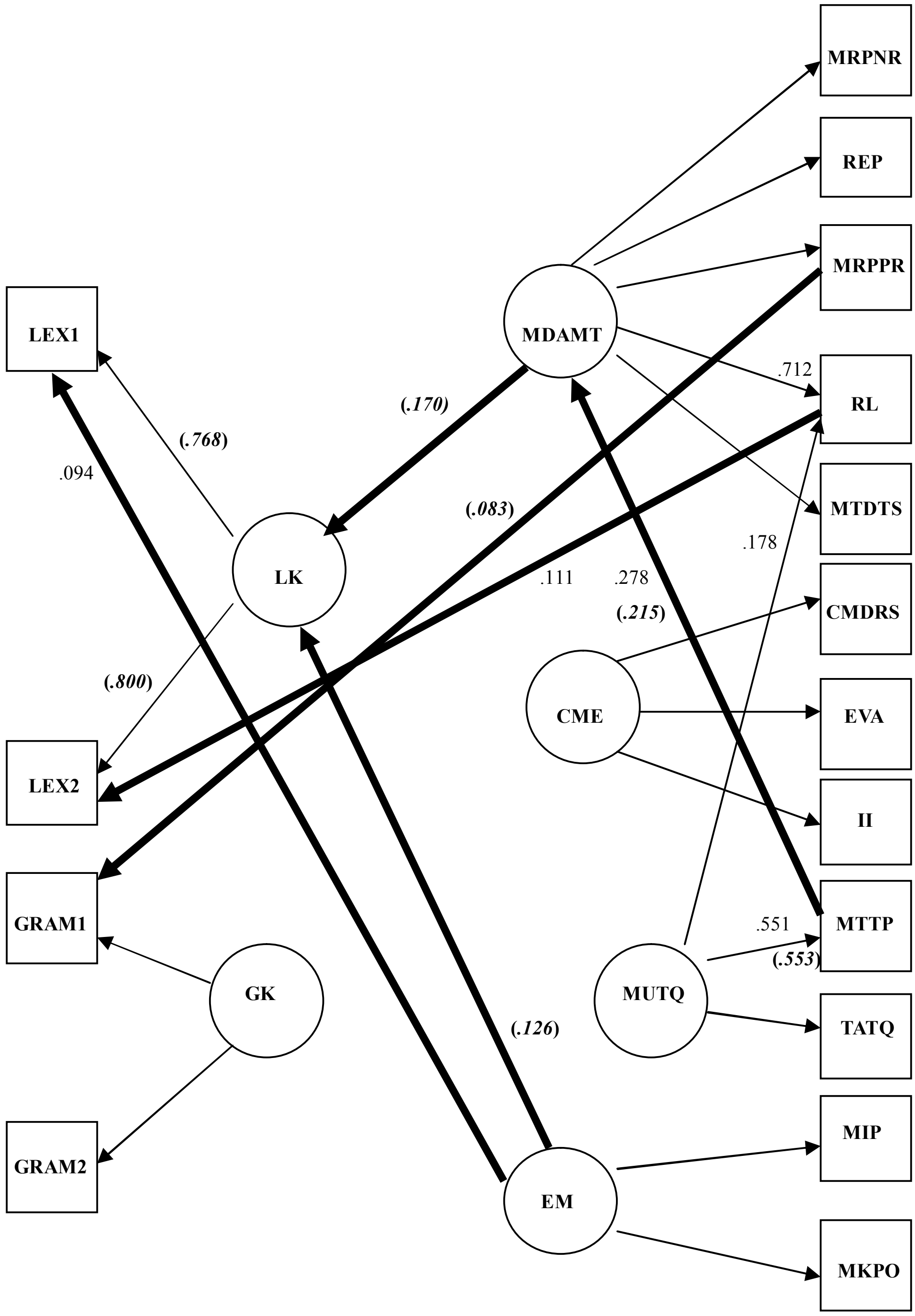
In this section, at first I examine an effect that English language knowledge (i.e., lexical knowledge and grammatical knowledge) yields on reading and test-taking strategy use (i.e., strategy use processes and strategy subgroups) for the HEA and the LEA groups and then shift to an impact that reading and test-taking strategy use has on English language knowledge.

Like that in the overall group model, effect paths that lexical knowledge or grammatical knowledge affects strategy use processes or strategy subgroups are observed in either group model. In this respect, the HEA and the LEA groups share some commonalities. For example, lexical knowledge (LK) exercises a trivial/weak, *direct*, positive influence on the *monitoring the reading process with positive results* (MRPPR) strategy subgroup in both groups, with values of .096 for the HEA group and .196 for the LEA group (see Figure 5.6). This cross-group commonality was further tested in the simultaneous group analysis. The HEA group also share an *indirect* effect with the LEA group – grammatical knowledge yields an effect on the MRPPR strategy subgroup via lexical knowledge. Finally, the GRAM2 test-item subgroup has a weak/trivial, direct, *negative* impact on the *taking advantage of test questions* (TATQ) strategy subgroup in the HEA and the LEA groups, with values of -.106 for the HEA group and -.097 for the LEA group (see Figure 5.6).



LK=Lexical knowledge; GK=Grammatical knowledge; LEX1 consists of ten test items of the vocabulary subtest; GRAM2 eight test items of the grammar subtest. MDAMT=Monitoring, directing attention and managing the test; REP=Repeating; MRPPR=Monitoring the reading process with positive results; II=Interacting with the input; TATQ=Taking advantage of test questions. Effects for the HEA group are not in the parenthesis, while those for the LEA group are in the parenthesis. ➡ =English language knowledge has an effect on strategy use.

Figure 5.6 The relationship between English language knowledge and reading and test-taking strategy use for the HEA group and the LEA group (part I): The separate group analysis



LK=Lexical knowledge; LEX1 consists of ten test items of the vocabulary subtest; LEX2 eleven test items of the vocabulary subtest; GRAM1 eight test items of the grammar subtest. MDAMT=Monitoring, directing attention and managing the test; EM=Evaluating and marking; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTTP=Monitoring the test-taking process. Effects for the HEA group are not in the parenthesis, while those for the LEA group are in the parenthesis. ➡ =Strategy use has an effect on English language knowledge.

Figure 5.7 The relationship between English language knowledge and reading and test-taking strategy use for the HEA group and the LEA group (part II): The separate group analysis

Table 5.4 Effects of English language knowledge on reading and test-taking strategy use and effects of reading and test-taking strategy use on English language knowledge for the HEA group and the LEA group: The separate group analysis

		Effects	
		The HEA Group	The LEA Group
Effects of ELK on RTSU	**[a] LK→MRPPR [f (g)]**	.096	.196
	[b] GRAM2→TATQ [f (g)]	-.106	-.097
	[b] GK→LK→MRPPR [f (g)]	.075	.116
	[b] GK→MDAMT [f]	.239	No
	[a] LEX1→TATQ [g]	No	-.098
	[b] GK→REP [g]	No	.104
	[b] GK→II [g]	No	.091
Effects of RTSU on ELK	[c] MUTQ→LEX2 (MUTQ→RL→LEX2 and MUTQ→MTTP→MDAMT→RL →LEX2 for the HEA group; MUTQ→ MTTP→MDAMT→LK→LEX2 for the LEA group) [f (g)]	.032	.016
	[c] EM→LEX1(EM→LEX1 for the HEA group; EM→LK→LEX1 for the LEA group) [f (g)]	.094	.097
	[c] RL→LEX2 [f]	.111	No
	[c] MDAMT→RL→LEX2 [f]	.079	No
	[c] MDAMT→LK [g]	No	.170
	[c] EM→LK [g]	No	.126
	[d] MRPPR→GRAM1 [g]	No	.083
	[c] MTTP→MDAMT→LK [g]	No	.037
	[c] MUTQ→MTTP→MDAMT→LK [g]	No	.020

Note. LK=Lexical knowledge; MRPPR=Monitoring the reading process with positive results; GRAM2 consists of eight test items of the grammar subtest; TATQ=Taking advantage of test questions; GK=Grammatical knowledge; MDAMT=Monitoring, directing attention and managing the test; LEX1 consists of ten test items of the vocabulary subtest; REP=Repeating; II=Interacting with the input; MUTQ=Monitoring and utilizing test questions; LEX2 consists of eleven test items of the vocabulary subtest; MTTP=Monitoring the test-taking process; EM=Evaluating and marking; RL=Retrieving-linking; GRAM1 consists of eight test items of the grammar subtest. The completely equivalent effect path shared by both groups is boldfaced. a=Lexical knowledge affects strategy use processes or strategy subgroups; b=Grammatical language affects strategy use processes or strategy subgroups. c=Strategy use processes or strategy subgroups affects lexical knowledge; d=Strategy use processes or strategy subgroups affects grammatical knowledge. “ f ” for the HEA group and “ g ” for the LEA group – English language knowledge affects strategy use and vice versa.

The HEA group vary from the LEA group in paths regarding lexical language or grammatical knowledge yielding an effect on strategy use processes or strategy subgroups. To illustrate, in the HEA group model, grammatical knowledge (GK) exerts a weak, positive effect on the *monitoring, directing attention and managing the test* (MDAMT) process, with a value of .239 (see Figure 5.6). By contrast, in the LEA group, the LEX1 test-item subgroup has a trivial, negative effect on the *taking advantage of test questions* strategy subgroup, with a value of -.098 (see Figure 5.6). The LEA group appear not to handle the processing in a well-balanced way that part of lexical knowledge is accessed to deploy taking-advantage-of-test-questions strategies. It could be argued that given the limitation of cognitive resources, gaining access to partial lexical knowledge to tap into test questions with an eye to promoting test performance exceed the LEA group's capacity. This may result in the LEX1 test-item subgroup being slightly detrimental to the *taking advantage of test questions* strategy subgroup.

Based on the results mentioned thus far, several implications are given. Firstly, students' English language knowledge, regardless of their English ability, is not always profitable to their strategy deployment. Further, both groups' English language knowledge has an impact on their strategy use in either a direct or an indirect way within this reading test-taking context. Either the HEA group or the LEA group, when taking this reading test, access their English language knowledge to a certain degree to invoke their strategies. Distinct from the HEA group, the LEA group's lexical knowledge has a negative effect on their strategy use partially due to their deficiency of lexical knowledge.

On the other hand, effect paths concerning strategy use processes or strategy subgroups impacting on lexical knowledge or grammatical knowledge are manifested within both group models. In this regard, there are variances between the groups. To illustrate, in the HEA group, the *evaluating and marking* (EM) process shows a trivial, *direct*, positive effect on the LEX1 test-item subgroup, with a value of .094 (see Figure 5.7). This result can be taken as an indication that for the HEA group assessing what has been processed and conducing marking is beneficial to their partial lexical knowledge access or vocabulary learning. In addition, the *retrieving-linking* (RL) strategy subgroup yields a weak, *direct*, positive effect on the LEX2 test-item subgroup, with a value of .111 (see Figure 5.7). The result suggests that the HEA group's employing retrieving and linking strategies promotes part of their lexical language knowledge access in the reading test-taking context.

Turning to the LEA group, the *monitoring, directing attention and managing the test* (MDAMT) and the *evaluating and marking* (EM) processes show a weak, *direct*, positive effect on lexical knowledge (LK), with values of .170 and .126 (see Figure 5.7). Such results indicates that the LEA group's employment of monitoring, repeating, retrieving-linking, managing-the-test and marking strategies contributes to their lexical knowledge access during this reading test. Aside from direct effects, indirect effects of strategy use on English language knowledge are present in the LEA group. For example, by means of the *monitoring, directing attention and managing the test* (MDAMT) process, the *monitoring the test-taking process* (MTTP) strategy subgroup has a trivial, *indirect* impact on lexical knowledge (LK), with a value of .037 (see Table 5.4).

The results stated so far suggest that students' strategy use, in spite of a difference in their English ability, exerts a trivial or weak, positive effect on their English language knowledge in a direct or indirect manner. Put differently, their strategy use makes divergent degrees of contributions to their English language knowledge in this test-taking context. Interestingly, the frequency that strategy use has an impact on English language knowledge in the LEA group is higher than that in the HEA group (7 vs. 4). This might suggest that the LEA group apply strategies to assist them in accessing English language knowledge more frequently than the HEA group.

To summarize, similar to that in the overall group, in the HEA and the LEA groups, English language knowledge has a direct or indirect impact on strategy use and vice versa. In addition, different from that in the HEA group, the relationship between strategy use and English language knowledge is more intricate in the LEA group. The HEA group's English language knowledge interacting with their strategy use differs from the LEA group's to some extent.

5.4.1.7 The relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

In this section, I first inspect an indirect effect that multiple-choice reading comprehension test performance (MC RCTP) receives from English language knowledge (ELK) and reading and test-taking strategy use (RTSU) for the HEA and the LEA groups. Then, I examine a total effect that ELK and RTSU show on MC RCTP.

A cross-group difference is pinpointed as the structure of the relationships among English language knowledge, strategy use, and reading test performance for the two levels are examined closely. Within the LEA group model, unexpectedly all strategy use processes and one strategy subgroup have an indirect impact on multiple-choice reading comprehension test performance (MC RCTP) by means of lexical knowledge (LK). More specifically, the *monitoring, directing attention and managing the test* (MDAMT), the *constructing the meaning and evaluating* (CME), the *monitoring and utilizing test questions* (MUTQ), and the *evaluating and marking* (EM) processes as well as the *monitoring the test-taking process* (MTTP) strategy subgroup all display trivial or weak, positive effects on MC RCTP. The values range from .016 to .131 (see Table 5.5). The result suggests that the LEA group's strategy use makes a trivial or weak indirect contribution to their reading test performance by means of lexical knowledge.

Table 5.5 Indirect effects of reading and test-taking strategy use on multiple-choice reading comprehension test performance through English language knowledge for the HEA group and the LEA group: The separate group analysis

	Effects	
	The HEA Group	The LEA Group
MDAMT→**LK**→MC RCTP	No	.131
CME→MDAMT→**LK**→MC RCTP	No	.046
MTTP→MDAMT→**LK**→MC RCTP	No	.028
MUTQ→MTTP→MDAMT→**LK**→MC RCTP	No	.016
EM→**LK**→MC RCTP	No	.097

Note. GK=Grammatical knowledge; MC RCTP=Multiple-choice reading comprehension test performance; LK=Lexical knowledge; MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; MTTP=Monitoring the test-taking process; EM=Evaluating and marking.

Different from the case in the LEA group, no strategy use processes or strategy subgroups, in the HEA group, exercise an influence on multiple-choice reading comprehension test performance (MC RCTP) by means of lexical knowledge (LK) or grammatical knowledge (GK). This comes as a surprise. I originally expected that strategy use processes or strategy subgroups, like the case in the LEA group, had indirect impacts on MC RCTP through LK or GK, given that these two groups shared the partial structure of the relationship amongst English language knowledge, strategy use and reading test performance, as mentioned in Sections 5.4.1.4, 5.4.1.5 and 5.4.1.6. However, this expectation, finally, fails to be observed in the full latent variable model for the HEA group. The HEA group's strategy employment may develop to a stage at which reliance upon English language knowledge is less demanded or English language knowledge can be accessed in an automatized way when strategies are invoked in this reading test context. Thus, that strategy deployment has an effect on reading test performance with English language knowledge as a mediator fails to be observed in the HEA group model.

A cross-group discrepancy is also reflected in indirect effects that English language knowledge exerts on reading test performance via strategy use. Similar to the case in the overall group, within the HEA group, grammatical knowledge has a trivial, indirect effect on multiple-choice reading comprehension test performance (MC RCTP) via the *monitoring, directing attention and managing the test* process, with a value of .034. The result suggests that the HEA group's grammatical knowledge, through their strategy use, makes a trivial, indirect contribution to their performance on the reading test. Conversely, no English language knowledge exerts an indirect influence on MC RCTP by means of strategy use within the LEA group. The result here, coupled with the aforementioned results, implies that the HEA group differ from the LEA group to some extent in (a) strategy use in concert with English language knowledge and (b) English language knowledge in combination with strategy utilization to boost their reading test performance.

With regard to total effects of English language knowledge (ELK) and reading and test-taking strategy use (RTSU) on multiple-choice reading comprehension test performance (MC RCTP), commonalities and differences are shown across English ability levels. As for commonalities, like that in the overall group, for the HEA and the LEA groups, ELK yields much more total effects on MC RCTP than RTSU. Specifically speaking, ELK, in both groups, yields a moderate or strong effect on MC RCTP, with values ranging from .464 to .857, while RTSU has a trivial or weak impact on MC RCTP, with values varying from .022 to .251 (see Table 5.6).

Although both English language knowledge (ELK) and reading and test-taking strategy use (RTSU) affect multiple-choice reading comprehension test performance (MC RCTP) across English ability levels, how they influence MC RCTP varies between the HEA and the LEA groups to a certain degree. To begin with, English language knowledge is centered on. In the HEA group, grammatical knowledge (GK) shows even more impacts on MC RCTP than lexical knowledge (LK) (.857 vs. .464). The finding suggests that grammatical knowledge, for the HEA group, promotes more their reading test performance than lexical knowledge in this test setting. However, the LEA group manifests a different case. LK and GK yields similar total effects on MC RCTP (.771

vs. .772). This implies that the LEA group's lexical and grammatical knowledge almost equally enhance their performance on the reading test within this multiple-choice reading test context. Moreover, compared with that in the HEA group, lexical knowledge has more total impacts on multiple-choice reading comprehension test performance in the LEA group (.464 vs. 771). The LEA group appear to count on lexical knowledge more than the HEA group do to tackle this reading test. This finding is consistent with that stated in Section 5.4.1.4 in which effects of English language knowledge (ELK) on multiple-choice reading comprehension test performance (MC RCTP) are limited to those observed within the relation between ELK and MC RCTP.

Table 5.6 Total effects of English language knowledge and reading and test-taking strategy use on multiple-choice reading comprehension test performance for the HEA group and the LEA group: The separate group analysis

		Effects	
		The HEA group	The LEA group
ELK	LK→MC RCTP	.464	.771
	GK→MC RCTP	.857	.772
RTSU	MDAMT→MC RCTP	.142	.251
	CME→MC RCTP	.047	.088
	MUTQ→MC RCTP	.022	-.121
	MTTP→MC RCTP	.040	.054
	EM→MC RCTP	No	.097

Note. ELK=English language knowledge; RTSU=Reading and test-taking strategy use; LK= Lexical knowledge; MC RCTP=Multiple-choice reading comprehension test performance; GK= Grammatical knowledge; MDAMT=Monitoring, directing attention and managing the test; CME= Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; MTTP=Monitoring the test-taking process; EM=Evaluating and marking.

With reference to total effects of reading and test-taking strategy use (RTSU) to multiple-choice reading comprehension test performance (MC RCTP), two commonalities are shared between these two groups. First of all, the effect of RTSU on MC RCTP is limited for both groups. In addition, the *monitoring, directing attention and managing the test* (MDAMT) process yields the most effects on MC RCTP than other strategy use processes and a strategy subgroup. These findings provide two implications. Firstly, students' strategy deployment, regardless of their English ability, can make small contributions to their reading test performance. Secondly, in this test-taking setting, both groups are able to appropriately and effectively invoke monitoring, repeating, retrieving-linking, and managing-the-test strategies covered by the *monitoring, directing attention and managing the test* process. Therefore, this strategy use process facilitates their reading test performance most.

When making a between-group comparison, I found something interesting. In comparison with those in the HEA group, total effects of reading and test-taking strategy use (RTSU) on multiple-choice reading comprehension test performance (MC RCTP), within the LEA group, overall are slightly stronger. Such a finding differs from that in Section 5.4.1.5, which makes sense seeing that in Section 5.4.1.5 the effects of RTSU on MC RCTP are limited to those within the relation between RTSU and MC RCTP. In this section, given the involvement of English language knowledge (ELK), a different scenario is expected regarding the impact of reading and test-taking strategy use on multiple-choice reading comprehension test performance across groups with divergent English ability. That the inclusion of English language knowledge leads to the strength of the effects of strategy use on reading test performance reversing between the groups demonstrates a close relationship between English language knowledge and strategy utilization in the multiple-choice reading test-taking context. This relationship makes a difference to the extent to which strategy use has an impact on reading test performance.

Then, why are total effects of strategy use in the LEA group overall slightly stronger that those in the HEA group? Two explanations are available. Firstly, the HEA group turn to other cognitive resources, rather than strategy use, to tackle this reading comprehension test. These cognitive resources are not examined in the current study. Secondly, there are other factors (e.g., test anxiety) adversely affecting the HEA group's strategy deployment in the reading test-taking situation. These factors are not involved and analyzed in the present model, but their existence lessens positive effects that strategy employment yields on multiple-choice reading comprehension test performance (MC RCTP). These two explanations are supported by the fact that the disturbance (i.e., D1 in Figure 5.4) related to MC RCTP is statistically significant at the .050 level ($p < .050$) in the HEA group model. This suggests that for the HEA group something else, rather than variables investigated and included in the current study, shows an effect on MC RCTP.

While total effects of strategy use on reading test performance in the LEA group, on the whole, are slightly larger than those in the HEA group, the LEA group's strategy deployment is not appropriate enough, given a negative total effect of the *monitoring and* *utilizing test questions* (MUTQ) process on multiple-choice reading comprehension test performance (MC RCTP), with a value of -.121 (see Table 5.6). The LEA group appear not to take advantage of strategies covered in the current study aptly and effectually to the extent that their strategy use can always be beneficial to their reading comprehension test performance.

When further inspecting total effects of strategy use on reading test performance, I found an unexpected result. The *evaluating and marking* (EM) process has a trivial effect on multiple-choice reading comprehension test performance in the LEA group, with a value of .097. This is different from that in Section 5.4.1.5 where no effect is manifested. The result here suggests that the LEA group's employing marking strategies finally makes a petty contribution to their reading test performance, which is distinct from the case in the HEA group – no contribution or counter-contribution is made.

To summarize, in this test-taking setting, the HEA and the LEA groups' English language knowledge has more total impacts on their reading test performance than strategy use. Their strategy use exerts a weak influence on their reading test performance. Despite a number of commonalities between the two groups, cross-group differences are present in the extent to which lexical knowledge and grammatical knowledge influence, and in strategy deployment affect reading test performance during the overall reading test-taking process. The HEA group's grammatical knowledge yields more effects on their reading test performance than lexical knowledge. By contrast, the LEA group's grammatical knowledge and lexical knowledge exercise a similar influence on their reading test performance. In addition, the LEA group's *monitoring and utilizing test questions* process shows a weak, *negative* effect on their reading test performance, whereas the HEA group's a trivial, *positive* effect. Finally, indirect effects of strategy use on multiple-choice reading comprehension test performance, by means of English language knowledge, are present in the LEA group, but absent in the HEA group. In the next section, the simultaneous group analysis will be addressed.

5.5 Simultaneously estimating the full latent variable models of the HEA group and of the LEA group

The separate group analyses show valuable information about full latent variable models for groups with different English ability. Despite the presence of cross-group discrepancies, the HEA and the LEA group models share a number of paths in all measurement models and an overall structural model. While cross-group similarities or differences in parameter estimates for equivalent paths and correlations manifested in both groups are pinpointed in the separate group analysis, this does not signify that the same result is produced as the two group models are estimated simultaneously. For example, the effect path that grammatical knowledge on multiple-choice reading comprehension test performance is shared by these two groups, and the effect of grammatical knowledge on multiple-choice reading comprehension test performance in the HEA group, is stronger than that in the LEA group (.459 vs. .315, see Table 5.2, p. 135) in the separate group analysis. However, we do not assure whether there is a significant difference in this path coefficient (i.e., the effect) across English ability levels. Accordingly, I performed the simultaneous group analysis to test whether the cross-group similarities or differences in parameter estimates for paths and correlations shared by the two levels still hold statistically.

5.5.1 The procedures for testing cross-group invariance

The models generated for the HEA and the LEA groups in the separate group analyses were estimated simultaneously, with equality constraints imposed on factor loadings in measurement models, and path coefficients and correlation coefficients in structural models. The current study adopted partial measurement invariance in testing for the equality of parameters. More specifically, not all parameters were constrained to be equivalent across the HEA and the LEA groups – the parameters without equality constraints encompassed (a) those pre-fixed with 1 for identification purposes, (b) those not present in both group models, and (c) those for factor variances, error variances or error correlations. Finally, twenty-four parameters were constrained to be invariant across these two groups.

The equality constraints were released one by one and produced models were competed with each other, with evaluations being made according to model fit indices and critical ratios for difference between parameters. More specifically, when an equality constraint was released, a new model was produced. This new model was evaluated based on whether model fit indices of this model were better than those of previous models and whether a critical ratio for difference for the parameter with the equality constraint just being released was not within ± 1.960 (i.e., $p < .050$). If both criteria were satisfied, then the released equality constraint would not be re-imposed on that parameter in the next analysis where another equality constraint of a parameter was released and another new model was generated. This demonstrated that the HEA group varied from the LEA group in the parameter of which the cross-group equality constraint was released. On the contrary, if either criterion was not met, then the released equality constraint would be re-imposed on that parameter in the following analysis. This revealed that the parameter of which the cross-group equality constraint was just released was invariant across these two groups. These procedures were carried out until all twenty-four cross-group equality constraints were examined one by one.

5.5.2 The results for cross-group tests of invariance

After twenty-four cross-group equality constraints were released one by one and produced models were inspected, finally the model with seventeen equality constraints was accepted (see Appendix 19 for the accepted model). Table 5.7 presents model fit indices for the model with twenty-four parameters freely estimated (i.e., the model with no equality constraints), the model with equality constraints on twenty-four parameters (i.e., the model with the most equality constraints), and the accepted model to justify the appropriateness for the acceptance of this model.

As indicated in Table 5.7, these three models exhibited fair goodness-of-fit – all the GFI, the AGFI, the CFI, and the TLI exceeded the cut-off value. Moreover, the statistics of the RMSEA were all below the .060 threshold. A comparison of these models showed that Model A performed best in the GFI (.970) and shared the same values with Model C in the AGFI (.954), the CFI (1.000), and the RMSEA (.000), suggesting that Model A might fit the collected data best. However, when several models are compared with one another, an inspection of the *Akaike's information criterion* (*AIC*) and the *Browne-Cudeck criterion* (*BCC*) indices is necessary – the smaller values for both indices symbolize the better fit of the postulated model (Byrne, 2001; Hu & Bentler, 1995). Table 5.7 shows that Model C displayed the smallest values for the AIC and the BCC among these models. Further, Model C exhibited satisfactory values for the GFI (.968), the AGFI (.954), the CFI (1.000), the TLI (1.002) and the RMSEA (.000). All of these statistics were better than those in Model B and were nearly as good as those in Model A. Hence, based on what has been discussed, Model C depicted the data best and was appropriate to be accepted as a model of choice.

Table 5.7 The model fit indices for the model with twenty-four parameters freely estimated, the model with equality constraints on twenty-four parameters and the model with equality constraints on seventeen parameters

Model fit indices	Levels of acceptable fit	Evaluation results		
		Model A	Model B	Model C
χ^2	Nonsignificant with the p-value > .050	Good (220.389 with p = .537 > .050)	Poor (294.094 with p = .021 < .050)	Good (234.148 with p = .594 > .050)
GFI	> .900	Very good (GFI = .970)	Very good (GFI = .960)	Very good (GFI = .968)
AGFI	> .900	Very good (AGFI = .954)	Good (AGFI = .945)	Very good (AGFI = .954)
CFI	> .950	Very good (CFI = 1.000)	Very good (CFI = .990)	Very good (CFI = 1.000)
TLI	> .950	Very good (TLI = 1.001)	Very good (TLI = .988)	Very good (TLI = 1.002)
RMSEA	< .060	Very good (RMSEA = .000)	Good (RMSEA = .016)	Very good (RMSEA = .000)
AIC	The smaller, the better.	(458.389)	(484.094)	The best (438.148)
BCC	The smaller, the better.	(471.460)	(494.529)	The best (449.351)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation; AIC=The Akaike's information criterion index; BCC=The Browne-Cudeck criterion index. Model A=The model with twenty-four parameters freely estimated; Model B=The model with equality constraints on twenty-four parameters; Model C=The model with equality constraints on seventeen parameters.

5.5.3 A close look at the results for cross-group tests of invariance

In this section, I first explain the results of cross-group invariance tests in general. Then, I examine and briefly discuss the results related to the measurement model as well as the structural model.

The results of cross-group invariance tests indicate that the equality constraints do not hold across the groups on seven parameters pertinent to the component structures of reading and test-taking strategy use and multiple-choice reading comprehension test performance, as well as the structural relationship amongst English language knowledge, strategy use and reading test performance. This reveals that there are differences between the HEA and the LEA groups on these parameters. Table 5.8 provides the seven parameters found not equivalent across the groups after cross-group invariance tests.

In the measurement model of reading and test-taking strategy use (RTSU), five parameters are found not to be equivalent across the two levels. The result suggests that how the reading and test-taking strategy use questionnaire functions varies to some extent across the groups with different English ability, although the two groups share the similar underlying component structure of the measurement model of RTSU.

Table 5.8 Cross-group tests of invariance

		The HEA group	The LEA group
		Factor loadings	
The measurement model of RTSU	MDAMT→REP	.468	.673
	MDAMT→MRPPR	.799	.585
	MDAMT→MTDTS	.460	.680
	MUTQ→RL	.180	.298
	EM→MKPO	.791	.922
The measurement model of MC RCTP	MC RCTP→InQ	.737	.608
		Effects	
The structural model of MC RCTP, ELK and RTSU	LK→MC RCTP	.505	.752

Note. RTSU=Reading and test-taking strategy use; MDAMT=Monitoring, directing attention and managing the test; REP=Repeating; MRPPR=Monitoring the reading process with positive results; MTDTS=Managing the test with the deployment of test-taking strategies; MUTQ=Monitoring and utilizing test questions; RL=Retrieving-linking; EM=Evaluating and marking; MKPO= Marking key points or options; MC RCTP=Multiple-choice reading comprehension test performance; InQ=Inferential questions; ELK=English language knowledge; LK=Lexical knowledge.

As indicated in the above table, the *repeating* (REP) and the *managing the test with the deployment of test-taking strategies* (MTDTS) strategy subgroups generate more loadings on the *monitoring, directing attention and managing the test* (MDAMT) process in the LEA group than in the HEA group (.673 vs. .468 for REP and .680 vs. 460 for MTDTS). This suggests that compared with the HEA group's utilization, the LEA group's deployment of repeating and managing-the-test strategies contributes more to their overseeing the reading process, directing attention to incomprehensible parts and administering the test well.

By contrast, the *monitoring the reading process with positive results* (MRPPR) strategy subgroup produces more loadings on the *monitoring, directing attention and managing the test* (MDAMT) process in the HEA group than in the LEA group (.799 vs. .585). Such a result manifests that the HEA group's use of partial monitoring strategies is more beneficial than the LEA group's employment to their supervising the reading process, channeling attention into comprehension breakdowns and managing the test adequately.

Moreover, the *marking key points or options* (MKPO) strategy subgroup yields more loadings on the *evaluating and marking* (EM) process in the LEA group than in the HEA group (.922 vs. .791). The result reveals that the LEA group's marking key points or options than the HEA group's enhances more the evaluating and marking process.

Finally, the *monitoring and utilizing test questions* (MUTQ) process receives more cross-loadings from the *retrieving-linking* (RL) strategy subgroup in the LEA group than in the HEA group (.298 vs. .180). The result indicates that the LEA group's employment of retrieving-linking strategies, in comparison with the HEA group's deployment, is more profitable to their overseeing the test-taking process and tapping into test questions in order to perform the test well.

All the results mentioned above suggest some discrepancies in strategy deployment between these two groups in the reading test-taking context, as other strategy-related studies in different contexts suggest (e.g., Nikolov, 2006; Oxford *et al*., 2004; Purpura, 1998b; 1999; Upton & Lee-Thompson, 2001). Further, the mean of the reading comprehension test for the HEA group was 10.969, while that for the LEA group was 9.242. In spite of the limited discrepancy, the HEA group outperformed the LEA group significantly in the reading comprehension test. Thus, the results stated above also imply that successful L2 readers vary from less successful L2 readers in part of their strategy employment.

On closely inspecting the five strategy subgroups not invariant across the two levels, I found interesting information. The three strategy subgroups, *retrieving-linking* (RL), *managing the test with the deployment of test-taking strategies* (MTDTS) and *marking key points or options* (MKPO), which produce more loadings in the LEA group than in the HEA group all more or less consist of strategies relevant to test-taking. For example, the RL strategy subgroup encompasses the strategy of *when I answered test questions, I tried to recall a part of the passage*; the MTDTS strategy subgroup *when I answered test questions, I tried to spend more time on difficult test questions*; the MKPO strategy subgroup *when I answered test questions, I tried to mark the differences among options*. This suggests that the LEA group, as sitting this reading comprehension test, appear more test-taking oriented than the HEA group.

With regard to the measurement model of multiple-choice reading comprehension test performance (MC RCTP), the parameter for the path between inferential questions

(InQ) and MC RCTP is found not equivalent across the groups. Such a result suggests that the multiple-choice reading comprehension test operates differently between the HEA and the LEA groups to a certain degree. As shown in Table 5.8, InQ yields more loadings on MC RCTP in the HEA group than in the LEA group (.737 vs. .608). This indicates that inferential questions assess multiple-choice reading comprehension test performance better in the HEA group than in the LEA group. This makes sense, as one can imagine that inferential questions require students to make sense of the passage thoroughly and to read between the lines in order to arrive at correct answers with more chances. The HEA group with more cognitive resources, no doubt, can outperform the LEA group in this type of test item. Then, it is no wonder that inferential questions measure the HEA group better than the LEA group.

Further, answering inferential questions entails students' piecing together information which may spread across the passage and then drawing inferences after digesting the input, which elicits more the top-down or the interactive reading. Then, the result noted above suggests that the HEA group conduct more the top-down or the interactive processing of reading than the LEA group in this reading comprehension test. This illustrates a difference in how to approach the reading task across groups with discrepant English ability.

Finally, as pointed out in the previous page, the HEA group performed slightly better on the reading comprehension test than the HEA group. Then, the aforementioned result also implies that successful L2 readers tend to deploy more the top-down reading, as a number of L2 reading studies suggest (e.g., Block, 1992; Hosenfeld, 1984; Yiğter *et al*., 2005).

Turning to the structural model, the equality constraint on one parameter does not hold across the groups. As can be seen in Table 5.8, lexical knowledge (LK) yields more effects on multiple-choice reading comprehension test performance (MC RCTP) in the LEA group than in the HEA group (.752 vs. .505), which confirms that pinpointed in Section 5.4.1.4. The result suggests that lexical knowledge makes more contributions to reading test performance in the LEA group than in the HEA group. In comparison to the HEA group, the LEA group seem to access and rely on lexical knowledge more greatly during this reading comprehension test. This might be because the LEA group more engage in tackling vocabulary which they are not familiar with than the HEA group. The explanation makes sense given the finding shown in Nikolov's (2006) study – low achievers utilize the strategy of "picking unknown vocabulary items" (p. 42) more frequently in the reading and writing test-taking course.

Finally, as mentioned in Section 5.4.1.5, the *monitoring, directing attention and managing the test* (MDAMT) process yields a positive effect on reading test performance in either case. The result of invariance tests shows that no cross-group significant difference is present in the effect for this path. This provides more robust evidence for the finding that the HEA group's and the LEA group's deployment of monitoring, repeating, retrieving-linking, managing-the-test strategies covered by the MDAMT process makes a similar direct contribution to their performance on the reading test.

In summary, reading and test-taking strategy use and multiple-choice reading comprehension test performance do not function in the same way across the two levels, because the hypothesis regarding cross-group equivalence on parameters of interest is not completely supported. Additionally, how the partial structure of the relationships among English language knowledge, strategy use, and reading test performance shared by the two levels performs varies to some extent between the groups.

5.6 Conclusion

This chapter is concerned with the multiple group analyses. The separate group analysis and the simultaneous group analysis were performed to pinpoint the commonalities and variations in the model regarding English language knowledge, strategy use, and reading test performance across English ability levels. As regards the measurement models, the separate group analysis and the simultaneous group analysis indicate that reading and test-taking strategy use and reading comprehension test performance do not operate in a fully equivalent manner between the two groups. However, English language knowledge does.

Turning to the structural model, the separate group analysis reveals that more interactions are present in the structural relationship among English language knowledge, strategy use, and reading test performance in the LEA group model than in the HEA group model. While both groups' English language knowledge and strategy use exert effects on their reading test performance, how they affect reading test performance is different across the groups. With regard to English language knowledge, the HEA group's grammatical knowledge yields more total effects on their reading test performance than lexical knowledge. In contrast, the LEA group's grammatical knowledge and lexical knowledge have similar impacts on their reading test performance. In addition, there are some cross-group variations in strategy use. To illustrate, the LEA group's strategy use overall shows more effects on their reading test performance than the HEA group's does. Also, some of the LEA group's strategy deployment has an influence on their reading test performance by means of English language knowledge. But the HEA group's does not. The LEA group's *evaluating and marking* process has a trivial, indirect, positive impact on their reading test performance, while the HEA group's displays *no* effect on their reading test performance. Finally, the LEA group's *monitoring and utilizing test questions* process has a weak, *negative* effect on their reading test performance; however, the HEA group's exerts a trivial, *positive* one.

With regard to the simultaneous group analysis, the result reveals the presence of a difference in the extent to which lexical knowledge yields an effect on multiple-choice reading comprehension test performance between the two levels. More specifically, lexical knowledge exercises more influences on multiple-choice reading comprehension test performance in the LEA group, compared with the case within the HEA group.

In the next chapter, I will discuss the major findings based on the research questions posed in the current study.

CHAPTER SIX
DISCUSSION

6.1 Introduction

The previous two chapters provide the results of analysis. This chapter concerns discussion of the major findings. Research questions are presented, the findings related to the questions are shown and answers to the questions are given. Comparisons are also made between the findings of the current study and those of previous studies.

This chapter is structured in the following order. To begin with, I discuss the major findings related to the relationship among students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance. Then, I shift to the major findings pertinent to the aforementioned relationship across English ability levels.

6.2 The relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance

The first research question addresses the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance. For this question, the current study shows the following findings.

Firstly, both English language knowledge and reading and test-taking strategy use are not single-dimensional constructs. English language knowledge is represented by two constructs: lexical knowledge and grammatical knowledge (see pp. 249-251 for details). Reading and test-taking strategy use is symbolized by four constructs: (a) the *monitoring, directing attention and managing the test* process; (b) the *constructing the meaning and evaluating* process; (c) the *monitoring and utilizing test questions* process; and (d) the *evaluating and marking* process (see pp. 254-256 for details).

Secondly, English language knowledge exercises an influence on multiple-choice reading comprehension test performance in direct and indirect ways. For example, lexical knowledge has a strong, *direct*, positive effect on reading test performance and grammatical knowledge a moderate, *direct*, positive impact (see Figure 4.7, p. 107 for details). Aside from the *direct* effect, grammatical knowledge also exerts a moderate, *indirect*, positive effect on reading test performance by means of lexical knowledge (see p.

107 for details) and a trivial, *indirect*, positive effect via a strategy use process (see p. 117 for details).

Thirdly, reading and test-taking strategy use also impacts on multiple-choice reading comprehension test performance in direct and indirect manners. For example, the *monitoring, directing attention and managing the test* process exercises a weak, *direct*, positive influence on reading test performance (see Figure 4.8, p. 109 for details). The *constructing the meaning and evaluating* process yields a trivial, *indirect* effect on reading test performance through the *monitoring, directing attention and managing the test* process (see p. 110 for details). In addition, the *constructing the meaning and evaluating* process shows a trivial, *indirect*, positive effect on reading test performance through English language knowledge, or English language knowledge and other strategy use process (see Table 4.3, p. 117 for details).

Finally, English language knowledge has a direct and indirect effect on strategy use. In one instance, lexical knowledge shows a weak, *direct*, positive effect on the *monitoring the reading process with positive results* strategy subgroup (see Table 4.2, p. 116 for details). Through lexical knowledge, grammatical knowledge has a weak, *indirect*, negative impact on the *taking advantage of test questions* strategy subgroup. Aside from strategy subgroups, English language knowledge also affects strategy use processes. An example for this is that grammatical knowledge shows a weak, direct, positive impact on the *monitoring, directing attention and managing the test* process (see Table 4.2, p. 116 for details). On the other hand, strategy use influences English language knowledge directly and indirectly. For example, the *constructing the meaning and evaluating* process displays a weak, *direct*, positive effect on grammatical knowledge. The *monitoring and utilizing test questions* process has a trivial, *indirect*, positive impact on lexical knowledge by means of other strategy use process (see Table 4.2, p. 116 for details).

The aforementioned findings indicate (a) no single-dimensional constructs underlying students' English language knowledge and their strategy use, and (b) potential relationships that take place among students' English language knowledge, strategy use, and their reading test performance in the entire test-taking process. Given these, we can conclude that the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their reading test performance is multi-directional, and at times subtle and interactive. Such a relationship provides useful information for the college entrance examination center and English language teachers in Taiwan. Caution should be taken when the score of the reading comprehension test is interpreted and a decision is made based on the score, since it represents the result that several students' cognitive resources (e.g., different types of language knowledge or discrepant forms of strategy use) interact with each other. In addition, more light is cast on the effect of strategy use on students' reading test performance. When the effect of strategy use is referred to, it should be kept in mind that the effect is more than what is believed to be – it covers the outcome that strategy use interacts with students' other cognitive resources (e.g., language knowledge), not just the effect of strategy use itself. When implementing strategy instruction, English language teachers in Taiwan ought to take into consideration students' other cognitive resources (e.g., language knowledge), since they are related to the impact that students' strategy use has on reading test performance.

In addition, the relation stated above provides an implication for the definition of strategic competence in Bachman and Palmer's (1996) model of language ability. To explain, given the complicated and subtle relationship among students' English language knowledge, strategy use and their reading test performance, it is reasonable to argue that there should be something to administer and organize students' (test-takers') intricate language knowledge access and strategy use in a reading test-taking setting. However, strategic competence in Bachman and Palmer's (ibid.) model of language ability just subsumes as a set of metacognitive strategies such as goal-setting, planning and assessment. As Purpura's (1997; 1998b; 1999) studies imply, even students with metacognitive strategies will draw upon their strategies so inappropriately that their test performance is inhibited. Phakiti (2003: 48) also remarks that "the use of a valid strategy implies neither an understanding in the need to use them, nor an awareness of the pitfalls of using a less adequate strategy". Accordingly, something more ought to be included in the construct of strategic competence. As suggested in the literature on metacognition and reading (e.g., Baker & Brown, 1984; Jiménez, *et al.*, 1996; Mokhtari & Reichard, 2004; Phakiti, 2003), metacognitive awareness plays a critical role in a reading process. In the reading domain, metacognitive awareness refers to readers' being able to think about their reading processes. Such awareness should be involved in strategic competence – test-takers' being capable of thinking about their test-taking processes. With the inclusion of metacognitive awareness as its construct, strategic competence can appropriately match its function as a high-level executive mechanism to orchestrate all cognitive activities in a test-taking context. Given the inclusion of metacognitive awareness, we might term strategic competence as metacognitive competence so that its name can correspond to its construct more precisely.

Previous qualitative-oriented studies on reading or test-taking strategies generally offer three aspects of substantive information. First of all, these studies have shown that strategy use exerts an effect on performance on L2 reading tasks or tests (e.g., Cohen & Upton, 2006; 2007; Hosenfeld, 1984; Rupp, Ferne, & Choi, 2006; Yang, 2002; 2006). This is based on the evidence that readers invoke strategies to deal with their comprehension breakdowns and the breakdowns are solved by their employment of strategies. Or test-takers turn to strategies to arrive at a plausible answer. Secondly, these research works demonstrate that L2 language knowledge or proficiency has a relation with strategy use during the L2 reading (e.g., McLeod & McLaughlin, 1986; Upton & Lee-Thompson, 2001; Yamashita, 2002). This is predicated on the finding that participants with different L2 language knowledge or proficiency deploy strategies in a not completely equivalent way. Thirdly, these studies suggest that L2 language knowledge or proficiency may have an impact on strategy use in the L2 reading course (e.g., Block, 1992; Clarke, 1979; 1980; Yang, 2002; 2006). This is grounded on the finding that readers with low L2 language knowledge or proficiency fail to deploy some strategies appropriately or effectively in their L2 reading. Or sometimes readers are aware of their incomprehensible parts, but they can not deal with them or cope with them properly with their strategy employment.

However, because of the limitations of data analysis methods, these qualitative-oriented studies fail to provide a picture of the relationship among English language knowledge, strategy use, and reading test performance within a single modeling framework in a way that effect paths among variables are shown. More specifically, they are not able to indicate the possible paths that constructs underlying English language knowledge and those underlying strategy use directly or indirectly impact on those underlying reading test performance. Furthermore, they do not reveal the potential paths that constructs underlying English language knowledge directly or indirectly interact with those underlying strategy use. Due to these drawbacks, these qualitative-oriented research works manifest little information on the multi-directional and sometimes interactive linkage amongst students' L2 language knowledge, strategy use, and their reading test performance. Then, there is a high likelihood that the construct of the effect of strategy use on performance on reading tests or tasks is simplified. A fallacy may arise that the effect of strategy use just represents the impact of strategy employment itself and nothing else.

Purpura's (1997; 1999) quantitative studies possibly are the first work to examine the relationship between strategy use and performance on L2 tests (reading comprehension, cloze, vocabulary and grammar tests) with the use of an SEM approach. His studies provide empirical evidence for (a) multi-dimensional constructs of strategy use and L2 test performance; (b) strategy use having a direct or indirect effect on L2 test performance. He concluded that the relationship between strategy use and L2 test performance was complex and occasionally subtle. Although valuable information is shown, three limitations are present in his studies. Firstly, his participants' not referring to given tasks, when they filled in the questionnaire, might lead to the collected data being somewhat unreliable. Secondly, in Purpura's (ibid.) studies, participants' L1 was heterogeneous, and so was their course level. Such variances may impact on participants' strategy use and L2 test performance. The findings regarding the relationship between strategy employment and L2 test performance could be impinged upon. Finally, his studies concentrated on the relationship between metacognitive and cognitive strategy use, and performance on L2 tests; thus, limited information was offered on the relationship among L2 language knowledge, strategy use, and L2 test performance, although lexico-grammatical ability was measured and included in his model.

The current study, distinct from Purpura's (1997; 1999), centers on the relation amongst English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance. Participants' strategy use is elicited with the presence of a given task. In addition, participants' L1 and their course level are homogeneous. With an SEM approach, the present study demonstrates the multi-directional, and sometimes subtle and interactive relationship among English language knowledge, strategy use, and reading test performance. Here the significance of the current study is illustrated. Specifically speaking, the current study offers more empirical evidence for Purpura's finding regarding the intricate relationship between strategy use and L2 test performance. Further, unlike Purpura's, the present study shows that the relation among English language knowledge, strategy use, and reading test performance is sometimes interactive – the interactive relationship is present between English language knowledge and strategy use. This is addressed further in Section 6.2.3.

6.2.1 The contributions of English language knowledge and strategy use to reading test performance

The first sub-question of the first research question concerns whether students' English language knowledge and strategy use contribute to their reading test performance. For this sub-question, the positive answer is given. To illustrate, the current study indicates that English language knowledge and strategy use, generally, exert weak to strong, *positive* total effects on reading comprehension test performance, despite the presence of a trivial, *negative*, total effect of some strategy use (see Table 4.4, p. 118 for details). The finding provides empirical evidence for Bachman's (1990) factors that influence test scores (for these factors, see pp. 36-37) and Bachman and Palmer's (1996) model of language ability (for this model, see p. 38). More specifically, language knowledge, strategic competence, and personal attributes/characteristics (strategy use is focused on in the present study) are components which impact upon test performance. In the following subsections, the contributions of English language knowledge to reading test performance are focused on at first.

6.2.1.1 The contributions of English language knowledge to reading test performance

As noted above, the present study has revealed that English language knowledge (i.e., lexical knowledge and grammatical knowledge) yields a strong, positive effect on multiple-choice reading comprehension test performance. In other words, students' English language knowledge heavily contributes to how well they perform the reading comprehension test, which concurs with that shown in other studies concerning L2 reading, L1-L2 reading or L2 assessment (e.g., Bernhardt & Kamil, 1995; Kobayashi, 2002; Purpura, 1997; 1998b; 1999; Shiotsu & Weir, 2007; Taillefer, 1996; Usó-Juan, 2006). Such a finding suggests that in this reading test-taking context, Taiwanese students heavily rest on their English language knowledge (i.e., lexical knowledge and grammatical knowledge) to deal with the test. Their English language knowledge appears a determinant for reading texts in English smoothly or performing an EFL reading test well. In order to process texts in English or test questions in a reading test, they need a certain amount of English language knowledge as a departure. That is, it is necessary for them to cross a basic language threshold regarding the amount of English language knowledge to succeed in performing an EFL reading task/test to some extent, as several

L2 reading researchers remark (e.g., Alderson, 1984; Devine, 1988; Usó-Juan, 2006; Yang, 2002; 2006).

The finding mentioned above also can be taken as an indication that Taiwanese senior high school students are equipped with certain levels of lexical knowledge and grammatical knowledge to rely upon, so that both types of language knowledge can greatly promote their performance on the reading comprehension test of which reading passages and test items were drawn from the reading comprehension subtest of the English component at the Senior High Academic Ability Examination (SHAAE). This is fair information for English language teachers at the senior high school level – their teaching is worthy and in a sense conducive to students' achievements.

In this study, English language knowledge consists of lexical knowledge and grammatical knowledge. Students' lexical knowledge and grammatical knowledge respectively promote their reading test performance *strongly*, with effects of .664 and .846. However, among other studies in which SEM is applied, van Gelderen *et al.*'s (2004) and Shiotsu and Weir's (2007) studies show a different picture of the contributions of lexical knowledge and grammatical knowledge to EFL reading test performance.

In van Gelderen *et al.*'s (2004) research work in which the relationship amongst linguistic knowledge, processing speed and metacognitive knowledge in L1 and L2 were examined. The data was collected from high school students. They found that students' L2 vocabulary knowledge *weakly* enhanced their L2 reading test performance, with an effect of .26. However, students' L2 grammatical knowledge neither facilitated nor inhibited their L2 reading test performance, because it showed no significant effect on L2 reading test performance.

Shiotsu and Weir (2007) reported three studies regarding the relative contributions of lexical knowledge and grammatical knowledge to EFL reading test performance. University students served as participants. In their report, students' grammatical knowledge *moderately* or *strongly* facilitated their reading test performance, with effects of .47, .61 and .64 in the three studies. With regard to students' vocabulary knowledge, *weak* or *moderate* contributions were made to their reading test performance, with effects of .42, .34 and .25.

Compared with those in the current study, the contributions made by students' lexical knowledge and grammatical knowledge to their reading test performance in van Gelderen *et al.*'s (2004) and Shiotsu and Weir's (2007) studies are smaller. This can be explained by several reasons such as participants' current lexical and grammatical knowledge, the way that lexical knowledge and grammatical knowledge are measured, participants' other cognitive resources except their lexical knowledge or grammatical knowledge, the way that EFL reading test performance is assessed, variables involved in the SEM model and so forth. No matter which one or ones, we can conclude that in L2/EFL reading, the contributions of lexical knowledge and grammatical knowledge to reading test performance is not fixed, but relative and subject to shift.

Within English language knowledge, the current study shows that grammatical knowledge exercises more positive influences on multiple-choice reading comprehension test performance than lexical knowledge (see Table 4.4, p. 118 for details). This suggests that students' grammatical knowledge is more beneficial to their reading test performance than lexical knowledge, which agrees with that in Shiotsu and Weir's (2007) study. In other words, when taking this multiple-choice reading test, students rely more on grammatical knowledge than lexical knowledge to tackle the test.

Notice that the reason why grammatical knowledge made a great contribution to reading test performance rests on the involvement of an *indirect*, positive effect that grammatical knowledge yields on reading test performance by means of lexical knowledge. This indirect effect is absent in Shiotsu and Weir's (2007) study. The presence of this indirect effect in the present study is because the effect path that grammatical knowledge impacts upon lexical knowledge is captured in the accepted SEM model.

The effect of grammatical knowledge on lexical knowledge observed in the current study partially results from how lexical knowledge is measured. As suggested by Read (2000), it is indeterminate to make a lucid distinction between lexical knowledge and other language knowledge in vocabulary tests. In the current study, the vocabulary subtest constitutes the sentence completion section and the definition matching section (see Appendix 1 for these test sections). Both sections require students to process a short sentence, and then to choose a correct answer. In order to minimize the influence that grammatical knowledge had on the vocabulary subtest, the sentence structure of test questions and options was simplified as much as possible. However, when sitting the vocabulary subtest, students in the sentence processing course still accessed grammatical knowledge to some extent to help them decide the meaning of the sentence. As a result, the casual effect between grammatical knowledge and lexical knowledge is captured in the full latent variable model pertinent to the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance.

Turning to Shiotsu and Weir's (2007) study, students were given two academic passages with several blanks. They filled in the blanks with words provided from a given word bank. Then, students needed to process the sentences with blanks embedded with in order to find appropriate words for blanks. It followed that they ought to access grammatical knowledge more or less to facilitate their processing during this course. When constructing the relation among lexical knowledge, grammatical knowledge and reading test performance in an SEM model, Shiotsu and Weir should have hypothesized that grammatical knowledge had an effect on lexical knowledge. However, they did not. They postulated a correlational relationship between grammatical knowledge and lexical knowledge. Then, their model should have shown poor goodness-of-fit. On the contrary, their model exhibited fair goodness-of-fit. This is attributed to the limited number of components investigated in their postulated model: only three (lexical knowledge, grammatical knowledge and reading test performance). Thereby, the degrees of freedom[8] were the same between the model where the correlational relationship between lexical knowledge and grammatical knowledge was hypothesized and the model where the causal relationship between lexical knowledge and grammatical knowledge was postulated. Both models featured appropriate goodness-of-fit. They chose the former one as their accepted model. This might be because the result obtained in this way is more easily interpreted to answer their research question – the relative contributions of lexical knowledge and grammatical knowledge to reading test performance. This illustrates one of the drawbacks of an SEM approach – we cannot tell whether the accepted model is a really true model that profiles the relationship of variables of interest. Additional studies with different measures to assess lexical knowledge and grammatical knowledge and the involvement of other variables are needed to shed more light on the relation between lexical knowledge and grammatical knowledge in L2 reading. Then, a more understanding will be given – whether it is better to treat lexical knowledge and grammatical knowledge as two separate aspects of language knowledge or as a combination, that is, lexcogrammar, as shown in Purpura's (1999) work.

[8] The degrees of freedom refer to the differences between the number of distinct sample moments and the number of distinct parameters to be estimated.

6.2.1.2 The contributions of strategy use to reading test performance

While previous qualitative-oriented studies have provided evidence for students' strategy use having an impact on their performance on reading tasks or tests (e.g., Block, 1986; Cohen & Upton, 2006; 2007; Hosenfeld, 1984; Yang, 2002; 2006), limited information is shown on the effect strength. What they can illustrate is that students surmount obstacles to their lack of comprehension by making use of strategies, and this obstacle-overcoming may facilitate their further comprehension of a passage or their reaching possible answers. With the application of SEM, the current study shows that students' strategy use, although not all, has a *weak*, positive effect on reading test performance (see Table 4.4, p. 118 for details). This suggests that students' strategy use promotes their reading test performance to some extent; however, the contributions are limited. The finding gives more empirical evidence for what Alexander *et al*.'s (1998) claim and several strategy-related studies (e.g., Anderson, 1991; Cohen & Upton, 2006; 2007; Phakiti, 2008; Purpura, 1997; 1999; Yang, 2002; 2006) indicate or suggest – the facilitative nature of strategies.

Among strategies of which students' deployment fosters reading test performance, their employment of the strategy group containing monitoring, repeating, retrieving-linking, and managing-the-test strategies *directly* promotes their reading test performance. This is based on the result that the *monitoring, directing attention and managing the test* (MDAMT) process consisting of these strategies has a direct, positive effect on multiple-choice reading comprehension test performance (MC RCTP) (see Figure 4.8, p. 109). The finding suggests that these students are able to deploy these strategies appropriately, so that a *direct* contribution can be made to their reading test performance. In this reading test, driven by a goal, they are aware of what they are going to do first. They often check their comprehension of and modify their hypothesis about what they read. When not getting a grip on the input, they probably repeat the incomprehensible parts. They also retrieve their comprehended parts of a passage or their cognitive resources and link them with what they are processing. Finally, they make an effort to understand test questions appropriately, pay attention to test time, and spend more time on challenging questions.

Interestingly, in addition to direct contributions, students' strategy use is profitable to their reading test performance indirectly. To illustrate, students' deployment of the strategy group covering constructing-the-meaning and evaluating strategies, and that encompassing monitoring-the-test-taking-process and taking-advantage-of-test-questions strategies, indirectly promotes their reading comprehension test performance via their use of the strategy group including monitoring, repeating, retrieving-linking and managing-the-test strategies. This is based on the result that the *constructing the meaning and evaluating* (CME) process and the *monitoring and utilizing test questions* (MUTQ) process exercise indirect, positive influences on multiple-choice reading comprehension test performance (MC RCTP) by means of the *monitoring, directing attention and managing the test* (MDAMT) process (see Figure 4.8, p. 109). In this reading test-taking setting, these EFL students are so strategic that they are able to deploy strategies in combination with other strategies so as to contribute to their reading test performance. The finding concurs partially with several strategy-related research works (e.g., Cohen & Upton, 2006; 2007; Nikolov, 2006; Purpura, 1997; 1998b; 1999; Yang, 2006).

Moreover, the aforementioned finding suggests that the strategy group consisting of monitoring, repeating, retrieving-linking and managing-the-test strategies, for these students, plays a predominant role in this reading comprehension test. Without it as a mediator, some of their strategy use fails to contribute to their reading comprehension test performance. For these students, their deployment of these strategies appears to have developed to a stage at which they are able to manipulate these strategies appropriately and flexibly so that they can utilize other strategies adequately with the group of these strategies as a basis. Then, it is reasonable to argue that for students what matters is not whether they employ a particular strategy or a set of strategies with the same functions, or use it or them appropriately, but how well they can deploy a group of strategies with different functions (e.g., identifying incomprehension, repeating and evaluating). This is similar to what several strategy researchers (e.g., Anderson, 1991; 2005; Chamot, 2005; Cohen; 1998b; Macaro, 2004; 2006) have pointed out and also gives an implication for English teachers in Taiwan. When strategy instruction is implemented, the focal point ought to be on teaching students to deploy strategies with diverse functions in a combination way.

Importantly, while students' strategy use directly or indirectly enhances their reading test performance, not all does. Students' deployment of the strategy group encompassing monitoring-the-test-taking-process and taking-advantage-of-test-questions strategies *directly* inhibits their reading test performance. This is based on the result that the *monitoring and utilizing test questions* (MUTQ) process covering these strategies yields a weak, *negative*, direct effect on multiple-choice reading comprehension test performance (MC RCTP) (see Figure 4.8, p. 109). The finding lends support to the notion that the deployment of strategies does not necessarily facilitate task performance, as

Carrell (1992) alludes to and other strategy-related studies reveal or imply (e.g., Nevo, 1989; Padron & Waxman, 1988; Purpura, 1997; 1999; Sarig, 1987; Yang, 2006). The finding suggests that these EFL students' strategy use might not develop to a stage where they are capable of appropriately employing all their own strategies, so that their strategy use can always be beneficial to their performance on this reading test. They may need strategy instruction to improve their strategy use and then further optimize their reading test performance.

As noted in the last paragraph on pages 169-170, students' deployment of the strategy group containing monitoring-the-test-taking-process and taking-advantage-of-test-questions strategies is also *indirectly beneficial* to their reading test performance through their use of the other strategy group. The finding, coupled with that mentioned in the last paragraph on page 170, shows the complexity of the employment of these strategies. They suggest that it is necessary for students, with their own cognitive resources as a basis, to flexibly tap into strategies to tackle tasks encountered if they intend to optimize their task performance with the reliance upon strategy deployment, as several strategy researchers implies (e.g., Purpura, 1997; 1999; Yang, 2006). Without employing strategies in an adjustable way, students' strategy use possibly inhibits, not facilitates, their overall performance on designated tasks. This implies the importance of metacognitive awareness in the test-taking process.

In addition, the finding that students' deployment of monitoring-the-test-taking-process and taking-advantage-of-test-questions strategies promotes or inhibits their reading test performance suggests students' question-answering orientation during this reading test, given that most of these strategies capitalize on test questions as clues. An example for these strategies is *when I took the test, I tried to use clues from test questions to decide whether to read a particular part of the passage*. These intermediate-beginning or intermediate EFL students have some understanding of how to take advantage of test questions to grapple with this reading comprehension test, while the result may be unsatisfactory. They appear to view this reading test as a problem-solving task, as implied in other test-taking strategy research works (e.g., Cohen & Upton, 2006; 2007, Farr *et al.*, 1990; Rupp *et al.*, 2006). This makes sense, since it is a multiple-choice reading test with test questions and alternatives, not a regular reading task, with which students deal.

A close examination of monitoring, retrieving-linking, evaluating and taking-advantage-of-test-questions strategies involved in strategy groups of which students' employment facilitates or inhibits their reading test performance shows that these strategies more or less feature metacognitive components. For instance, a strategy of *during the reading process, I was aware that I did not understand a part of the passage*, with a monitoring component, is included in monitoring strategies. A strategy of *when I read the passage, I tried to identify the important and the less important parts of the passage*, with an evaluating element, is contained by evaluating strategies. The finding indicates that students' metacognitive awareness is involved in their reading test performance. They appear metacognitvely aware of their reading test-taking process to a certain degree. Further, the finding gives partial evidence for Rogers and Bateson's model of expert test-takers' test-taking behavior (1991; 1994) in which monitoring carries great weight in the test-taking process.

From the aforementioned finding, coupled with the finding that students' strategy use contributes to their reading test performance directly or indirectly, two implications are available. Firstly, these students are strategic to some extent when sitting L2 reading tests. This perhaps is not so surprising given that "strategies are mandatory (essential) for academic development" (Alexander *et al.*, 1998: 131). These EFL students' scores for the senior high entrance examination must cross a certain threshold, and then they can enter the senior high schools where the current study was undertaken. Immersed in such academic-oriented environments, these students probably are more aware that they need to perform well on their academic studies in order to attend prestigious universities after graduation. This need can lead them to approach a given task in a strategic fashion with a view to completing it successfully or maximizing the possibility of completing it successfully. Secondly, metacognitive awareness is influential in reading or test-taking processes, as work by a number of researchers reveals or implies (e.g., Anderson, 1991; Phakiti, 2003; Purpura, 1997; 1999; Sheorey & Mokhtari, 2001; Yang, 2006; Yang & Zhang 2002). Without metacognitive awareness, students fall short of invoking strategies in an appropriate, effective, and flexible manner, which may lead to their strategy use not contributing to their performance on a given task. This is valuable information for English language teachers in Taiwan. When strategy instruction is implemented, a certain amount of attention needs to be given to activate or develop students' metacognitive awareness.

To conclude, the current study presents several findings regarding strategy deployment. Firstly, strategy use yields an effect on how well students perform a reading test. Secondly, strategies are employed in a combination way to contribute to students' reading test performance. Thirdly, strategy deployment does not always promote performance on a reading test. Finally, metacognitive awareness exerts an influence on reading test performance. While such findings have been suggested in the qualitative studies (e.g., Block, 1986; 1992; Cohen & Upton, 2006; 2007; Jiménez, 1996; Nikolov, 2006; Yang, 2006), this quantitative study provides more empirical evidence for these by looking at the effect that strategy use has on reading test performance with the application of structural equation modeling.

6.2.1.3 The relative contributions of English language knowledge and strategy use to reading test performance

The second part of the first sub-question comprised by the first research question concerns the relative contributions of English language knowledge (i.e., lexical knowledge and grammatical knowledge) and strategy use to reading test performance. The SEM analysis result shows that lexical knowledge and grammatical knowledge respectively yield a *strong*, *positive* effect on reading test performance in the case in which the relationship between these two types of English language knowledge and reading test performance is focused on (see Figure 4.7, p. 107 for details). Similar results are manifested as total effects of lexical knowledge and grammatical knowledge on reading test performance in the overall SEM model are examined (see Table 4.4, p. 118 for details). However, strategy deployment exerts a *trivial* or *weak*, *positive* effect or a *weak*, *negative* effect on reading test performance when the relation between strategy use and reading test performance is centered on (see Figure 4.8, p. 109 and p. 110 for details). Similar results are revealed when total effects of strategy employment on reading test performance in the entire SEM model are inspected (see Table 4.4, p. 118 for details). Based on these results, it seems reasonable to conclude that students' English language knowledge, compared with strategy use, contributes more to their reading test performance. These students count on their English language knowledge more heavily than their strategy use to deal with the multiple-choice reading comprehension test.

The finding that English language knowledge, in comparison with strategy employment, more promotes EFL reading test performance is of importance for two reasons. Firstly, from an SEM perspective, the finding lends additional support to the notion that in L2 reading L2 language knowledge carries more weight and makes more contributions than strategy use, as implied by several L1-L2 reading research (e.g., Bernhardt & Kamil, 1995; Bossers, 1991; Lee & Schallert, 1997; Taillefer, 1996; Yamashita, 2002) and by Yang's (2006) reading strategy study.

Secondly, the finding suggests that the key to facilitating students' reading test performance rests on English language knowledge rather than strategy use, given that strategy use does not make so many contributions to reading test performance as expected. This offers a critical implication to English language teachers in Taiwan who intend to incorporate strategy instruction into regular English language classes. Strategies, rather than as an elixir of students' poor reading performance in test-taking or non-test-taking setting, are just alternative resources that students rely upon and invoke when they need to. Teachers need to reflect on how and when to implement their strategy instruction if they aim to optimize students' employment of these resources.

It is worth noting that Purpura's (1997; 1999) studies in which an SEM approach was conducted gave similar evidence regarding the relative contributions of English language knowledge and strategy use to reading test performance. In Purpura's research works, cognitive and metacognitive strategy use, compared with lexico-grammatical ability, made smaller contributions to reading test performance. More specifically, lexico-grammatical ability *very strongly* promoted reading test performance, with an effect of as high as .985. Strategy use weakly or moderately enhanced reading test performance, with a value ranging from .095 to .458. However, in Purpura's studies, a flaw is present – a nonsignificant path for strategy use having an effect on lexico-grammatical ability was retained in his model in order to obtain fair goodness-of-fit of his model. In doing so, the yielded results were influenced and the findings were questionable. With a similar finding as that in Purpura's studies, the present study confirms Purpura's (ibid.) evidence pertinent to the relative contributions of English language knowledge and strategy use to reading test performance.

However, the finding here is tentative and more extensive research works are still needed. As several strategy researchers have suggested (e.g., Anderson, 1991; Grenfell & Harris, 1999; Macaro, 2006; Oxford *et al*., 2004; Phakiti, 2003; Rupp *et al*., 2006), strategy use varies with several factors internal or external to users, such as users' declarative knowledge (know what) and procedural knowledge (know how) about strategies, the extent to which users can deploy certain strategies effectively, users' language proficiency, the attributes of given tasks and so forth. These variations impact on the contributions made by strategy use to performance on given tasks. An additional study can be carried out in which participants receive strategy use training to see whether English language knowledge still functions as a stronger contributor to EFL reading test performance than strategy use. The result can provide an implication for whether strategy use merits more emphasis being placed on in English language classes where usually several linguistic components need to be covered but available time is limited.

6.2.2 A language threshold for some strategy use to contribute to reading test performance

The second sub-question of the first research question concerns the presence of a language threshold for Taiwanese senior high school students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance. For this sub-question, the positive answer is given since the current study reveals that some of students' strategy use exercises trivial or weak, indirect, positive effects on their reading test performance by means of English language knowledge (see Table 4.3, p. 117 for details). In other words, some of students' strategy deployment entails their English language knowledge as a mediator and then the strategy use can be beneficial to their reading test performance in this test-taking setting. A language threshold is present for students' employing some reading and test-taking strategies to contribute to their reading test. The finding lends support to one of strategy features submitted by Macaro (2004) and the implication given by a number of L1-L2 reading studies (e.g., Bernhardt & Kamil, 1995; Bossers, 1991; Clarke, 1980; Lee & Schallert, 1997; Taillefer, 1996; Walter, 2004; Yamashita, 2002) that the deployment of strategies necessitates a certain amount of language knowledge if the strategy employment is intended to make a contribution to task performance.

In most L1-L2 reading studies (e.g., Bernhardt & Kamil, 1995; Bossers, 1991; Lee & Schallert, 1997; Taillefer, 1996), it is L1 reading ability that is concentrated on and assessed by L1 reading tests. Then, what is measured is L1 reading ability including L1 language knowledge, not strategies themselves. Because of this shortcoming, these studies fail to provide appropriate empirical evidence for the issue – whether there is a language threshold for employing some reading or test-taking strategies to promote performance on a reading task.

Different from previous L1-L2 reading studies, Yamashita (2002), with the use of think-aloud procedures, elicited strategies that university students deployed respectively in L1 reading and L2 reading and then compared their strategy employment between L1 reading and L2 reading among students with different L1 and L2 ability. Based on the discrepancies in some reading strategies that students invoked in L1 and L2 reading, she concluded that in L2 reading a language threshold was present for deploying some strategies (i.e. local and global reading strategies) which she called *language dependent strategies*. Although her evidence was based on strategies, rather than L1 reading ability, it did not provide robust evidence for the aforementioned issue. This is because in her study the number of participants is limited: twelve. In addition, more exactly, her evidence is for a language threshold for transferring L1 reading and test-taking strategies to L2 reading.

Purpura's (1997; 1999) quantitative research works indicated that strategy use exercised an indirect influence on reading ability via lexico-grammatical ability. This appears to give evidence for the presence of a language threshold for the deployment of some strategies to facilitate L2 test performance. However, such evidence is somewhat questionable. To explain, his studies focused on the relationship between strategy use and performance on L2 tests. Both reading ability and lexico-grammatical ability in his model were L2 test performance. Then, he should have hypothesized that strategy use had an impact respectively on reading ability and on lexico-grammatical ability, and reading ability was correlated with lexico-grammatical ability. However, he postulated lexico-grammatical ability had a direct effect on reading ability. This did not make sense since both were L2 test performance. Doing so resulted in the fact that there was something wrong with the validity of his postulated model to some extent.

From an SEM perspective, the current study provides more evidence for a language threshold for employing some reading and test-taking strategies to promote L2/EFL reading performance. With an eye to employing some strategies to contribute to L2/EFL reading performance, students need to be equipped with a certain amount of L2 or English language knowledge. Deficiency of L2 or English language knowledge short-circuits their deployment of some of strategies, even though they possess the strategies and are aware of invoking them. This highlights the importance of L2 or English language knowledge in strategy deployment and offers an implication to strategy instruction in an L2/EFL context. That is, students' L2/English language knowledge should be taken into careful consideration when strategy instruction is conducted since students' current L2/English language knowledge has something to do with whether utilization of some strategies has a positive impact upon performance on a given task. The presence of the effect further influences students' willingness to deploy certain strategies and the effectiveness of strategy instruction.

6.2.3 The relationship between English language knowledge and strategy use in a reading test

The third sub-question of the first research question is concerned with the relationship between students' English language knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test. For this sub-question, the current study reveals that students' English language knowledge exercises an influence on their strategy use and vice versa. More specifically, on the one hand, students' lexical knowledge and grammatical knowledge have a weak, direct, positive or negative impact on their strategy use (see Table 4.2, p. 116 for details). Apart from *direct* effects, their grammatical knowledge also yields a weak, *indirect*, negative effect on their strategy employment by means of lexical knowledge or lexical knowledge and other strategy deployment (see Table 4.2, p. 116 for details). On the other hand, students' strategy use displays a weak, *direct*, positive effect on their lexical knowledge and grammatical knowledge (see Table 4.2, p. 116 for details). In addition to *direct* effects, some of their strategy deployment also exerts a trivial, *indirect* influence on their lexical knowledge through other strategy employment (see Table 4.2, p. 116 for details). These results lead to a conclusion that within this reading test, students' English language knowledge interacts with their strategy use to a certain degree.

Rather than a fixed causal relationship, a temporary causal relationship or an interactive relationship exists between language knowledge and strategy use in the reading test-taking context. Driven by a given task, students access their English language knowledge in order to promote their strategy deployment. Sometimes they not simply get access to English language knowledge but also rely on some of their strategy use so as to enhance other strategy deployment. On the other hand, they make use of their strategies with a view to fostering their access to or learning of English language knowledge. In addition to employing strategies in an isolation way, they invoke strategies in a combination fashion to facilitate their accessing English language knowledge. In the reading test-taking setting, students' access to English language knowledge and their strategy use appear to intertwine with each other to some extent.

Two points are worth noting. First of all, the current study indicates the interactive relationship between English language knowledge and strategy use. A close examination of the strategy use that has an effect on English language knowledge and vice versa indicates that some strategies feature metacognitive components (e.g., monitoring or evaluating). For example, the *monitoring, directing attention and managing the test* process affects lexical knowledge, whilst grammatical knowledge influences this strategy use process (see MDAMT→LK; GK→ MDAMT on Table 4.2, p. 116). This strategy use process consists of strategies with monitoring components. Illustrations for these strategies are *during the reading process, I was aware that I understood a part of the passage* and *when I read the passage, I was aware of the difficulty of the passage*. It follows that the present study provides partial evidence for the interactive relationship between language knowledge and strategic competence (defined as a set of metacognitive strategies), as described in Bachman and Palmer's (1996) model of language ability.

Secondly, in the current study strategy use always yields a positive effect on English language knowledge, while English language knowledge does not always (see Table 4.2, p. 116 for details). The finding that strategy employment exercises a positive influence on English language knowledge supports the notion that strategy use contributes to language knowledge access or learning in the L2 reading context, as previous strategy-related studies[9] imply (e.g., Fraser, 1999; Kern, 1989). The finding also suggests that these students have good command of some strategies to the extent that their deployment of these strategies always enhances their access to or learning of English language knowledge. This is useful information for English language teachers in Taiwan, given that a general understanding is provided that Taiwanese senior high school students are strategic to a certain degree within the L2 reading test-taking setting. With such an understanding, when intending to implement strategy instruction, teachers can reflect on how to take advantage of students' current knowledge about strategies to improve students' strategy employment and further their performance on a reading test.

On the other hand, English language knowledge has a positive or a negative impact on strategy use. In other words, students' English language knowledge promotes or inhibits their strategy use. More specifically, students' lexical knowledge exercises a *positive* influence on their deployment of partial monitoring strategies (see LK→MRPPR on Table 4.2, p. 116). On the contrary, students' lexical knowledge has a *negative* impact on their use of managing-the-test strategies and taking-advantage-of-test-questions strategies (see LK→MTDTS; LK→TATQ on Table 4.2, p. 116). Similarly, students' grammatical knowledge also shows a *negative* effect their employment of taking-advantage-of-test-questions strategies (see GK→TATQ on Table 4.2, p. 116).

---

[9] These studies suggest that the employment of strategies instructed has a positive effect on vocabulary learning or vocabulary inferencing ability within the L2 reading setting.

What is stated above makes an implication. Whether students' English language knowledge impacts on their strategy use positively or negatively may be related to whether the deployment of a strategy subgroup requires linguistic processing to a great extent, that is, students need to heavily access English language knowledge for this strategy use. Due to the great requirement, students' accessing English language knowledge goes well with their deploying a certain strategy subgroup (e.g., a strategy subgroup of monitoring strategies related to comprehension-checking). Thereby, English language knowledge positively influences the deployment of such a strategy subgroup.

By contrast, the employment of some strategy subgroups (e.g., the use of a strategy subgroup of managing-the-test strategies) requires linguistic processing less heavily. Rather, the deployment of these strategy subgroups demands other processing greatly, such as evaluating processing. Such processing loads students with some cognitive loads and so does accessing English language knowledge. Therefore, students' accessing English language knowledge for the use of these strategy subgroups exceeds students' capacity, given the limitations of their ability. It follows that English language knowledge adversely affects the employment of these strategy subgroups. A further study merits being conducted in which strategies are grouped in a categorical way to provide insights into how different types of English language knowledge interact with discrepant forms of strategy groups in EFL reading context. Then, a clear picture can be provided of what types of strategy groups require linguistic processing greatly. Such information is helpful in strategy instruction. When implementing strategy instruction, teachers will be aware of what strategies entail linguistic processing and what strategies do not. Then, they can adjust their strategy instruction with students' language ability. The situation will be avoided that they teach students a set of strategies which necessitate linguistic processing to a certain degree to promote students' performance on a reading test, when students' language ability is still limited.

With more light on the relation between language knowledge and strategy use being cast, the current study demonstrates its significance in three aspects. Firstly, the present study provides more empirical evidence for the notion that L2 language knowledge impacts upon strategy use in the L2 context. Although the previous research works suggest that L2 language knowledge or L2 proficiency yields an effect on strategy use (e.g., Anderson, 1991; Clarke, 1980; McLeod & McLaughlin, 1986; Upton & Lee-Thompson, 2001; Yamashita, 2002), their evidence is "soft" and indirect. In these studies, participants were divided into two or more groups according to their L2 proficiency. By making between-group comparisons in strategy use, they found some cross-group differences in strategy use. Then, a conclusion is drawn that L2 language knowledge or L2 proficiency impacts on strategy use. However, such a conclusion appears questionable given that their evidence at best illustrates that L2 language knowledge or L2 proficiency is related to strategy use. The current study here provides strong and direct evidence for the notion that language knowledge exerts a positive or negative effect on strategy employment in a reading test-taking setting with effect paths manifested in the model.

Secondly, the present study also gives more evidence for the notion that strategy use has an effect on L2 language knowledge in the L2 context. The finding is slightly different from that in Purpura's (1997; 1999) studies in which SEM is adopted. In Purpura's (ibid.) studies, strategy use exerted either *positive* or *negative* effects on lexico-grammatical ability. By contrast, in the current study, strategy use yields only a *positive* effect on lexical knowledge or grammatical knowledge. This discrepancy between his studies and the present study is partially attributable to participants involved[10] and how lexical knowledge and grammatical knowledge were measured[11]. To illustrate, strategy use is subject to users and tasks (Grenfell & Harris, 1999; Nevo, 1989; Rupp *et al.*, 2006). Language learners with divergent language or strategy resources, as encountering tasks with different difficulty levels, might invoke strategies to discrepant degrees. Now that there are variations in participants and in how lexical knowledge and grammatical knowledge were assessed in Purpura's (ibid.) studies and the present study, strategy use varies to some extent across studies. Then, the slightly different finding regarding effects of strategy use on L2 language knowledge between his studies and the current study makes sense.

[10] In Purpura's (1997; 1999) studies, participants came from three countries: Czech Republic, Spain and Turkey. Their L1 was heterogeneous. Their English proficiency ranged from high beginning to proficiency. Participants consisted of high school students, university students and others not students. In the current study, participants were third-graders of senior high school students in Taiwan. Their L1 was homogeneous. They had learned English at least for five years.

[11] In Purpura's (1997; 1999) studies, lexico-grammatical ability was measured by word formation, sentence formation, vocabulary, and grammar tests. According to Purpura (1999), the word formation test was to measure "test-takers' ability to use English morphology to transform the root of a word into a related word form according to how the word is used in a sentecne" (p. 55). An example for the word formation test is that "He gave me a ____ of nuts and raisins (hand)". The sentence formation test was to measure "test-takers' ability to generate synonymous sentences" (p. 55). An example for the word formation test is that "I expect that he will get there by lunchtime. → I expect him ___________ ." The vocabulary and grammar test was to measure "test-takers' ability in the use of grammatical rules and constraints, semantic sets and collocations, and phrasal verbs" (p. 54). For example, "After the deaths of her parents the girl was ____ by her grandparents (A) brought up (B) grown up (C) taken up (D) given up." As for how lexical knowledge and grammatical knowledge were assessed in the current study, see pp. 62-63 and Appendix 1.

Finally, the previous strategy-related studies (e.g., Anderson, 1991; Purpura, 1997; 1999; Upton & Lee-Thompson, 2001; Yamashita, 2002), due to analytic methods utilized or the purpose of the research, only provide evidence for either an effect that L2 language knowledge/L2 proficiency yields on strategy use or an impact that strategy employment has on L2 language knowledge/L2 proficiency. With the use of SEM, the current study reveals effects that L2 language knowledge exerts on strategy use and that strategy use yields on L2 language knowledge in a single modeling framework, which is distinct from other related studies. Clearly, the current study overcomes some limitations of the previous studies.

6.3 The relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance across English ability levels

The second research question asks whether there is a difference in the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance across English ability levels. For this question, the answer is positive. The relationship among students' English language knowledge, strategy use, and their reading test performance is not completely the same between the HEA group and the LEA group based on the evidence from two sources. Firstly, the components underlying the relationship amongst students' English language knowledge, strategy use, and their reading test performance operate differently to some extent across the groups. Secondly, the structural relationship among students' English language knowledge, strategy use, and their reading test performance is not fully the same between the two levels.

More specifically, in terms of the components composing the aforementioned relation, strategy use and reading test performance do not perform in an entirely equivalent manner across English ability levels. First of all, strategy use is focused on.

With regard to strategy use, the variation between the HEA group and the LEA group is reflected in two aspects. Firstly, the HEA group do not share all the component structure of reading and test-taking strategy use with the LEA group. Secondly, while the partial component structure is shared across the groups, how it works varies to a certain degree.

For the first aspect, the HEA group do not share with the LEA group in three cross-component loadings in the component structure of reading and test-taking strategy use. In the HEA group model, the *managing the test with the deployment of test-taking strategies* strategy subgroup generates a cross-loading on the *monitoring and utilizing test questions* process (see MUTQ→MTDTS on Figure 5.2, p. 129 for details). Distinct from the HEA group, within the LEA group, the *monitoring the reading process with positive results* strategy subgroup produces a cross-loading on the *constructing the meaning and evaluating* process (see CME→MRPPR on Figure 5.2, p. 129 for details). Further, the *managing the test with the deployment of test-taking strategies* strategy subgroup yields a cross-loading on the *evaluating and marking* process (see EM→MTDTS on Figure 5.2, p. 129 for details).

For the second aspect, the simultaneous group analysis result reveals that cross-group variances are present in five factor loadings in the component structure of reading and test-taking strategy use, shared by both groups. First of all, compared with that within the LEA group, the *monitoring the reading process with positive results* strategy subgroup yields more loadings on the *monitoring, directing attention and managing the test* process in the HEA group (see MDAMT→MRPPR on Table 5.8, p. 155 for details). By contrast, the *repeating* and the *managing the test with the deployment of test-taking strategies* strategy subgroups generate more loadings on the *monitoring, directing attention and managing the test* process in the LEA group than in the HEA group (see MDAMT→REP; MDAMT→MTDTS on Table 5.8, p. 155 for details). Additionally, the *monitoring and utilizing test questions* process receives more cross-loadings from the *retrieving-linking* strategy subgroup in the LEA group, in comparison with the case in the HEA group (see MUTQ→RL on Table 5.8, p. 155 for details). Finally, the *marking key points or options* strategy subgroup produces more loadings on the *evaluating and marking* process in the LEA group than within the HEA group (see EM→MKPO on Table 5.8, p. 155, for details).

What is mentioned above indicates cross-group variations in strategy employment across English ability levels, as suggested in previous strategy-related studies (e.g., Nikolov, 2006; Oxford *et al*., 2004; Purpura, 1998b; 1999; Upton & Lee-Thompson, 2001). Further, the aforementioned findings suggest that in the reading test-taking context, the HEA group appear to show an inclination to employ more monitoring strategies, while the LEA group seem to tend to invoke more test-taking and marking strategies.

As far as reading test performance is concerned, there is a variance in factor loadings in the component structure of multiple-choice reading comprehension test performance between the two groups. To explain, inferential questions produce more loadings on multiple-choice reading comprehension test performance in the HEA group model than in the LEA group model (see MC RCTP→InQ on Table 5.8, p. 155 for details).

In the light of the structural relationship amongst English language knowledge, strategy use and reading test performance, the HEA group differs from the LEA group to some extent. Such a discrepancy is manifested by two facets. Firstly, both groups do not share all the effect paths with each other. Secondly, even though effect paths are shared by the two groups, how they function is different across the groups.

For the first facet, the structure of the relationships among students' English language knowledge, strategy use, and their reading test performance in the LEA group model is slightly more complicated than that within the HEA group model. More interactions among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance are observed in the LEA group than those in the HEA group. For example, the path for the *monitoring and utilizing test questions* process directly affecting multiple-choice reading comprehension test performance is manifested in the LEA group model but not in the HEA group (see MUTQ →MC RCTP on Figure 5.5, p. 138). Additionally, the LEA group model captures the path that grammatical knowledge has an effect on the *repeating* strategy subgroup, but the HEA group model does not.

For the second facet, the simultaneous group analysis result shows that the path for lexical knowledge impacting upon multiple-choice reading comprehension test performance operates differently across these two groups, even though both group models capture the path. More specifically, the LEA groups' lexical knowledge exerts greater effects than the HEA group's does on reading test performance (see LK→MC RCTP on Table 5.8, p. 155 for details).

Based on what has been discussed thus far, it appears reasonable to conclude that the relationship among English language knowledge, strategy use, and reading test performance varies to some extent across English ability levels. How the LEA group access their English language knowledge and make use of strategies to tackle the multiple-choice reading comprehension test differs from how the HEA group do to some extent. Such information is useful in constructing a test-taking model regarding a multiple-choice test format to profile how Taiwanese senior high school students (test-takers) arrive at plausible answers with their English language knowledge and strategy use. Rather than one, maybe two models, one for the HEA group and the other for the

LEA group, are supposed to be formulated. Then, test results of multiple-choice reading tests can be interpreted more precisely and more "clearly communicated to test-takers and educational decision-makers" (Rupp *et al*., 2006: 470).

The present study shows a similar finding as Purpura's (1998b; 1999) that the relationship among English language knowledge, strategy use and reading test performance in the LEA group is slightly more complex than that in the HEA group. In this respect, the current study indicates its significance. To explain, within Purpura's studies, the results of the reading, vocabulary, and grammar tests, all of which were included in parameter estimation, were adopted to separate his participants into two groups. The current study as mentioned in Section 5.2 utilizes a different way to divide participants into groups and presents the finding similar to that in Purpura's studies. It follows that the current study gives more empirical evidence to the finding in Purpura's studies in spite of a discrepancy in what the current study and his studies concentrate on.

6.3.1 The contributions of English language knowledge and strategy use to reading test performance across English ability levels

The first sub-question of the second research question concerns whether there is a difference in students' English language knowledge and strategy use contributing to their reading test performance across English ability levels. For this sub-question, the answer is positive and negative. For the positive one, similar to the entire group, both groups' English language knowledge and strategy use are conducive to their performance on the reading test to a certain degree. For the negative one, there are cross-group variations in the contributions that different types of English language knowledge and of strategy use made to reading test performance. In the following subsections, the contributions of English language knowledge to reading test performance are centered on first.

6.3.1.1 The contributions of English language knowledge to reading test performance across English ability levels

The current study reveals that within the HEA group lexical knowledge yields a *moderate*, positive effect and grammatical knowledge has a *strong*, positive impact on multiple-choice reading comprehension test performance (see Table 5.6, p. 149 for details). In the LEA group, both lexical knowledge and grammatical knowledge exert *strong*, positive effects on multiple-choice reading comprehension test performance (see Table 5.6, p. 149 for details). These results lead to a conclusion that both groups' English language knowledge (i.e., lexical knowledge and grammatical knowledge) promotes their reading test performance. The finding concurs with that in Shiotsu and Weir's (2007) study and partially with Purpura's (1998b; 1999) studies.

However, cross-group differences are present in the extent to which different types of English language knowledge foster reading test performance. With respect to within-group comparisons, as mentioned above, lexical knowledge within the HEA group exercises a *moderate*, positive influence on multiple-choice reading comprehension test performance, while grammatical knowledge has a *strong*, positive one. The HEA group's grammatical knowledge contributes more to their reading test performance than lexical knowledge. The finding suggests that the HEA group appear to rest more on grammatical knowledge than lexical knowledge to tackle this reading test. On the other hand, as noted above, lexical knowledge and grammatical knowledge in the LEA group, yield *strong*, positive effects on multiple-choice reading comprehension test performance. The LEA group's lexical knowledge and grammatical knowledge make similar contributions to their performance on the reading test. The finding indicates that the LEA group seem to rely on grammatical knowledge as heavily as lexical knowledge when taking this reading test.

As far as between-group comparisons, the simultaneous group analysis result manifests that the LEA group's lexical knowledge than the HEA group's yields more effects on multiple-choice reading comprehension test performance (see Table 5.8, p. 155 for details). To put it another way, the LEA group's lexical knowledge, in comparison with the case in the HEA group, is more beneficial to their reading test performance. The finding implies that the LEA group draw upon lexical knowledge more than the HEA group to deal with this reading test. It can be argued that the LEA group, in comparison with the HEA group, appear to encounter more lexical problems or to conduct more local reading in their test-taking process, which leads them to rest on lexical knowledge more. If this is the case, English language teachers in Taiwan need to put more effort into teaching the LEA groups how to improve their lexical knowledge, figure out the meanings of unknown words from context, and read in a global way.

It is worth noting that the finding regarding the contributions of lexical knowledge and grammatical knowledge across English ability groups in the current study slightly differs from those in Shiotsu and Weir's (2007) study. In their study, within the HEA group grammatical knowledge displayed a *moderate*, positive effect on reading test performance, with a value of .50, whilst lexical knowledge showed a *weak*, positive one, with a value of .19. In the LEA group, grammatical knowledge had a *strong*, positive influence on reading test performance, with a value of .62, whereas lexical knowledge exercised a *weak*, positive one, with a value of .26. Their results indicate that no matter in the HEA group or the LEA group, grammatical knowledge was more profitable to reading test performance than lexical knowledge. In contrast, the current study manifests that grammatical knowledge facilitates reading test performance more than lexical knowledge in the HEA group, while grammatical knowledge and lexical knowledge make similar contributions to reading test performance in the LEA group. The discrepancy in these findings between the current study and Shiotsu and Weir's (2007) study can be explained by the following.

In the present study, the LEA group's English ability might be not good enough to take the reading test of which test items and reading passages were drawn from the Senior High Academic Ability Examination. For them, tackling this reading test is probably challenging; they encounter unfamiliar words to the extent which they need to access lexical knowledge greatly to work out their meanings. Such processing, for them, is in a controlled way, so that more effects are observed in the model. On the contrary, in Shiotsu and Weir's study, the LEA group were EFL college students whose English ability should be at a certain level, although they were labeled as the LEA group. Reading tests administered, for them, could not be demanding. Then, they were able to tackle the reading tests to the extent which they did not need to access lexical knowledge heavily. Part of their accessing lexical knowledge processing in the test-taking context was in an automatized fashion. Consequently, fewer effects were manifested in their model. Nonetheless, this is just an assumption, since other factors such as how reading test performance, lexical knowledge and grammatical knowledge are measured also impact upon the finding. Further research in which reading test performance, lexical knowledge and grammatical knowledge are assessed in a way different from that in the current study and Shiotsu and Weir's (2007) study is needed to provide more evidence for the relative contributions of lexical knowledge and grammatical knowledge to EFL reading test performance across English ability levels. The finding can serve as a reference point for English language teachers to adjust their teaching when they aim to improve students' reading test performance and students' English ability is different. Is more emphasis placed on lexical knowledge or grammatical knowledge, or both?

6.3.1.2 The contributions of strategy use to reading test performance across English ability levels

Similar to the entire group's strategy use, both groups' strategy deployment, although not all, displays a trivial or weak, positive effect on how well they perform the multiple-choice reading comprehension test (see Table 5.6, p. 149). Both groups' strategy deployment enhances their reading test performance to a certain degree, while the contributions of their strategy use are limited, compatible with the finding in Purpura's (1998b; 1999) research work.

Amongst all strategy use, either the HEA group's or the LEA group's deployment of the strategy group consisting of monitoring, repeating, retrieving-linking and managing-the-test strategies facilitates their reading test performance most. This is based on the result that the *monitoring, attention directing and managing the test* (MDAMT) process covering these strategies in both group models exerts the most positive effects on multiple-choice reading comprehension test performance (MC RCTP) (see Figure 5.5, p. 138 or Table 5.3, p. 139 for details). The finding suggests that these EFL students, regardless of their English ability, are able to tap into these strategies to the extent which their employment of these strategies can contribute to their reading test performance more than other strategy use.

In addition, both groups' employment of the strategy group subsuming constructing-the-meaning and evaluating strategies and that containing monitoring-the-test-taking-process and taking-advantage-of-test-questions strategies indirectly promotes their reading test performance through their use of the strategy group comprising monitoring, repeating, retrieving-linking and managing-the-test strategies. This is based on two results. Firstly, within both group models, the *constructing the meaning and evaluating* (CME) process yields a trivial, indirect, positive effect on multiple-choice reading comprehension test performance (MC RCTP) via the *monitoring, directing attention and managing the tes*t (MDAMT) process (see Figure 5.5, p. 138 and Table 5.3, p. 139 for details). Secondly, in both group models, the *monitoring and utilizing test questions* (MUTQ) process shows a trivial, indirect, positive effect on multiple-choice reading comprehension test performance (MC RCTP) via the *monitoring, directing attention and managing the test* (MDAMT) process (see Figure 5.5, p. 138 and Table 5.3, p. 139 for details). The finding indicates that in the reading test-taking setting, either group deploy their strategies not merely in an isolation way (a single strategy group) but in a combination fashion as well (a strategy group via the other). They make use of their strategies in an adjustable way. Given their flexible strategy use, it can be argued that both the HEA and the LEA groups are metacognitively aware of their test-taking course to some extent. This finding, coupled with that mentioned in the previous paragraph, provides an implication for English language teachers in Taiwan. That is, even the LEA group, like the HEA group, possess a set of strategies at their disposal and they are strategic L2 readers/test-takers in one sense. This information can function as a frame of reference when teachers intend to implement strategy instruction to improve the LEA groups' strategy employment and thereby their reading test performance.

However, the HEA group differs from the LEA group in the contributions of their strategy use to their reading test performance in two aspects. Firstly, some of the LEA group's strategy use inhibits their performance on the reading test. By contrast, all of the HEA group's strategy deployment promotes their reading test performance. More specifically, the LEA group's employment of the strategy group covering monitoring-the-test-taking-process and taking-advantage-of-test-questions strategies is directly detrimental to their reading test performance. This is based on the result that the *monitoring and utilizing test questions* (MUTQ) process has a weak, direct, negative impact on multiple-choice reading comprehension test performance (MC RCTP) (see Figure 5.5, p. 138 and Table 5.3, p. 139 for details). But this is absent in the HEA group. The finding implies that the LEA group's strategy use, compared with the HEA group's, is not appropriate to a certain degree, as other strategy research works suggest (e.g., Clarke, 1980; Cziko,1980; Nikolov, 2006). The LEA group appear to more need strategy instruction than the HEA group to improve their strategy use and further optimize their reading test performance.

Secondly, the LEA group's strategy use indirectly contributes to their reading test performance via their English language knowledge. However, the HEA group's strategy employment does not. This point will be discussed further in Section 6.3.2.

What has been discussed thus far leads to a conclusion that the contributions of both groups' strategy use to their reading test performance are not completely equivalent across these two groups in the EFL reading test-taking context. This provides more evidence for the notion that in the L2 context strategy use varies with users' L2 ability to a certain extent, as indicated or implied in previous strategy-related studies (e.g., Cziko, 1980; McLeod & McLaughlin, 1986; Nikolov, 2006; Oxford *et al*., 2004; Phakiti, 2003; Purpura, 1998b; 1999; Upton & Lee-Thompson, 2001) in test-taking or non-test-taking settings.

A point is worth noting. As mentioned in Section 6.3, the HEA group vary from the LEA group in the deployment of partial monitoring, repeating and managing-the-test strategies. These strategies are involved in a strategy group – the *monitoring, directing attention and managing the test* (MDAMT) process. Then, a cross-group difference should be present in the contributions of their use of this strategy group to their reading test performance. However, surprisingly, for the HEA and the LEA groups, the *monitoring, directing attention and managing the test* (MDAMT) process has a similar, direct, positive, effect on multiple-choice reading comprehension test performance (MC RCTP) (see Figure 5.5, p. 138 for the effect strength). In addition, the simultaneous group analysis result shows that no cross-group discrepancy is present in the effect strength. That is, there is no significant difference in the direct contribution of the strategy group containing partial monitoring, repeating and managing-the-test strategies to reading test performance across English ability levels. The findings indicate that although the HEA group and the LEA group invoke the same strategies to a certain divergent extent, as these strategies combine with other strategies to form a strategy group, this strategy group directly promotes reading test performance similarly between these two groups. This gives the following two implications.

Firstly, similar to what is outlined in Section 6.2.1.2, when it comes to strategy use, what matters is the appropriate employment of a strategy group consisting of several strategy subgroups with diverse functions for a given task. Strategy deployment can be thought of as "an orchestra. Rarely does an instrument sound good alone. However, when combined with other instruments, beautiful music results" (Anderson: 2005, 757). Similarly, only when strategies with discrepant functions are utilized simultaneously can the effect of strategy use on task performance be maximized. This is useful information for English language teachers. When strategy instruction is implemented, teaching students how to deploy strategies in a combination way should be the focus.

Secondly, something else (e.g., users' language knowledge or attributes of tasks) is involved in strategy use and interacts with it, as Bachman and Palmer's (1996) model of language ability suggests. It can be influential enough to make a variation in the contribution of the deployment of a strategy group to reading test performance so limited across English ability levels that the variation is rejected from a statistical perspective. Similar to what has been pointed out in Section 6.2, the construct of the effect of strategy use on a given task is multi-dimensional, not limited to strategy use of itself. As the effect of strategy use is addressed, it would rather be referred to as the effect of the consequence that strategy use interacts with users' other cognitive resources and attributes of tasks than as the effect of strategy use in itself.

6.3.1.3 The relative contributions of English language knowledge and strategy use to reading test performance across English ability levels

The second part of the first sub-question included in the second research question is concerned with whether the relative contributions of English language knowledge (i.e., lexical knowledge and grammatical knowledge) and strategy use to reading test performance differ across English ability levels. For this part of the sub-question, the answer is negative. The current study indicates that in the HEA group lexical knowledge shows a *moderate* and grammatical knowledge exerts a *strong*, positive effect on multiple-choice reading comprehension test performance, while strategy use yields a *trivial* or *weak* positive one (see Table 5.6, p. 149 for details). Within the LEA group, lexical knowledge and grammatical knowledge exercise *strong* influences on multiple-choice reading comprehension test performance, whereas strategy use has a *trivial* or *weak* one (see Table 5.6, p. 149 for details). Clearly, both groups' English language knowledge than strategy use exerts more positive effects on how well they perform the reading test. There is no difference in the relative contributions of English language knowledge and strategy use to reading test performance across English ability levels. These EFL students, no matter what their English ability level is, rely more on their English language knowledge than strategy employment to tackle the reading test. A critical implication is given for English language teachers in Taiwan, which is addressed as follows.

Given that the HEA group have already been equipped with a certain great amount of English language knowledge, teachers may plan to implement strategy instruction for them to promote their strategy use. Thereby, the HEA group will possess more resources to deal with reading tasks/tests encountered. However, a certain level of attention still ought to be given to develop and consolidate the HEA group's English language knowledge. Going all out for strategy instruction and taking little heed of the persistent accumulation of English language knowledge is the last thing to be observed. After all, for students regardless of their English ability, English language knowledge still plays a more dominant and influential role than strategy use in EFL reading.

The finding that English language knowledge than strategy employment contributes more to reading test performance is similar to that in Purpura's (1998b; 1999) and Yamashita's (2002) studies but distinct from that in Carrell's study (1991) on L1-L2 reading. In Carrell's (1991) study, for the group with low L2 ability, L2 language knowledge made more contributions to L2 reading test performance than L1 reading ability, while for the group with high L2 ability, the case reversed. The difference in the relative contributions of L2 language knowledge and L1 reading ability/strategies to L2 reading test performance across groups between her study and the present study can partially be attributable to the fact that in her studies it was L1 reading ability, rather than strategy employment, which was measured and analyzed.

As mentioned in Sections 1.3 and 6.2.2, limitations are present for Purpura's (1998b; 1999) and Yamashita's (2002) studies. Given that the current study shows a similar finding as theirs, obviously, the current study gives more empirical support for their findings, which highlights the significance of the present study.

6.3.2 A language threshold for some strategy use to contribute to reading test performance across English ability levels

The second sub-question of the second research question concerns whether there is a language threshold for students' deploying some reading and test-taking strategies to contribute to their multiple-choice reading comprehension test performance across English ability levels. For this sub-question, the answer is negative since the current study indicates different stories for groups with different English ability.

The present study indicates that strategy use in the LEA group yields a trivial or weak indirect, positive effect on multiple-choice reading comprehension test performance by means of English language knowledge (see Table 5.5, p. 147). That is, the LEA group's strategy deployment makes an indirect contribution to their reading test performance through English language knowledge in the test-taking setting. The LEA group get access to their English language knowledge for some of their strategy use and thereby the strategy employment indirectly boosts their performance on the reading test. A language threshold is present for the LEA group's employing some strategies to contribute to their reading test performance.

On the contrary, strategy employment within the HEA group has no indirect impact on multiple-choice reading comprehension test performance via English language knowledge (see Table 5.5, p. 147). In other words, the HEA group's strategy use neither indirectly facilitates nor inhibits their performance on the reading test by means of their English language knowledge. The HEA group might automatize the process that they turn to their English language knowledge for their strategy employment; thus, this process is not observed in their group. There appears to be an upper language threshold for employing some reading and test-taking strategies. The HEA group cross it, so that some of their strategy use entailing English language knowledge as a mediator is not captured.

The aforementioned findings provide more evidence for the cross-group discrepancy in the contributions of strategy use to reading test performance, as discussed in Section 6.3.1.2. This cross-group difference supplies English language teachers in Taiwan with an implication. When strategy instruction is integrated into regular English classes, students' English ability needs to be taken into account. If allowed, teachers may separate students into two groups according to their English ability. For the high English ability group, more strategy instruction can be given to enhance their strategy use. The focus can be on how to employ strategies in a combination manner and sophisticated strategies can be centered on (e.g., predicting the content of the following paragraph). As for the low English ability group, the focal point is on improving their English ability, although strategy instruction is implemented. Less sophisticated strategies are presented (e.g., using grammar rules to analyze a sentence). In this way, both groups can benefit from teaching.

Additionally, the findings here offer an implication to a language threshold for strategy deployment. There seems to be two language thresholds for strategy deployment in the L2 context: the lower one and the upper one. Crossing the lower one, then L2 students can employ some strategies to contribute to their task performance with the assistance of L2 language knowledge, just like the LEA group and the entire group in the current study. One the other hand, once crossing the upper one, the process that L2 students invoke some strategies with the reliance upon language knowledge to promote their task performance becomes automatized, just like the HEA group in the present study. With such an implication, the current study complements Ridgway's (1997) study in the existence of two language thresholds – in his study evidence for an upper language threshold for drawing on background knowledge in L2 reading was not found.

Notice that the findings noted earlier differ from those in Purpura's studies (1998b; 1999). In his studies, the HEA group's and the LEA group's strategy use facilitates or inhibits their reading test performance through lexico-grammatical ability. The difference is attributable to participants and tasks, given that strategy deployment is subject to users and tasks encountered (Grenfell & Harris, 1999; Nevo, 1989; Rupp *et al.*, 2006). In the current study, a multiple-choice reading comprehension test is administered to participants who are third-graders of senior high with homogeneous L1 and their course level. Within Purpura's studies, a reading comprehension test, a cloze test, a vocabulary test and a grammar test were given to participants with heterogeneous L1 and their course level (see Footnotes 10 and 11 on p. 180 for details). Now that there are discrepancies between the current study and his studies in participants and tasks given, the fact that the finding regarding strategy use via language knowledge promoting reading test performance across English ability levels is different is reasonable. However, as mentioned in 6.2.2, there is something wrong with his hypothesized model; thus, his finding is questionable regarding a language threshold for some strategy use to contribute to reading test performance in the L2 context across English ability levels.

6.3.3 The relationship between English language knowledge and strategy use in a reading test across English ability levels

The third sub-question of the second research question is concerned with whether there is a difference in the relationship between students' English language knowledge and their reading and test-taking strategy use in a multiple-choice reading comprehension test across English ability levels. For this sub-question, the answer is positive and negative since commonalities and variations are present between these two groups, which are addressed as follows.

The current study reveals that similar to the entire group, both the HEA and the LEA groups' English language knowledge interact with their strategy use within this reading test-taking context. Specifically speaking, on the one hand, both groups' English language knowledge shows a trivial or weak, positive or negative effect on their strategy use (see Table 5.4, p. 144 for details). On the other hand, these two groups' strategy deployment yields a trivial or weak, direct or indirect, positive effect on their English language knowledge (see Table 5.4, p. 144 for details).

While the interactive relation between English language knowledge and strategy use is observed across the groups, limited commonalities are shared across these two groups in effect paths concerning either English language knowledge impacting upon strategy use or vice versa – both groups share only three completely equivalent effect paths (see Table 5.4, p. 144). Additionally, the total number that English language knowledge influences strategy use and vice versa is eight for the HEA group, while thirteen for the LEA group (see Table 5.4, p. 144 for details). The LEA group's English language knowledge interacts with their strategy use more frequently, compared with the case in the HEA group. The findings indicate English language knowledge does not interact with strategy deployment in fully the same way across English ability groups in the reading test-taking situation. More specifically, the LEA group, in comparison with the HEA group, seem more actively to draw upon strategies to access language knowledge needed and to rely on language knowledge to employ strategies within this test-taking setting. This can be taken as an indication that the LEA group access their English language knowledge or deploy their strategies in a more controlled manner than do the HEA group. Then, English language teachers in Taiwan can think about how to help the LEA group improve their English language knowledge access and strategy deployment processes so that these two processes or part of them can be in an automatic manner. Perhaps, adding more English classes? Integrating strategy instruction into regular English classes? Instructing strategy use for a certain span of time? Providing more opportunities in class and out of class for the LEA group to practice what they have learned? When the LEA group can access their language knowledge or deploy strategies automatically on most occasions, they will have more spare ability to access their other cognitive resources (e.g., knowledge of subject matter) to deal with a designated task. Then, they will stand a better chance of performing a given test well.

Two points are worthy of noting. In both group models, among the paths that English language knowledge influences strategy use, the total number of the paths that grammatical knowledge affects strategy use is five (see Table 5.4, p. 144 for details). In contrast, the total number of the paths that lexical knowledge affects strategy use is only two (see Table 5.4, p. 144 for details). Grammatical knowledge exerts an effect on strategy use more frequently than lexical knowledge does. The finding implies that the HEA and the LEA groups appear to count on their grammatical knowledge more greatly than lexical knowledge to contribute to their strategy use. Their grammatical knowledge seems better than their lexical knowledge. This information gives an implication for English language teachers in Taiwan. It appears necessary for teachers to adjust their teaching focus – perhaps more emphasis is placed on augmenting students' lexical knowledge.

On the other hand, within these two group models, among the paths strategy use affects English language knowledge, the total number of the paths that strategy use influences grammatical knowledge is one (see Table 5.4, p. 144 for details). By contrast, the total number of the paths that strategy use influences lexical knowledge is eight (see Table 5.4, p. 144 for details). Strategy deployment has more positive impacts on lexical knowledge than on grammatical knowledge. The finding suggests that these two groups' strategy use enhances more their lexical knowledge access or learning than grammatical knowledge access or learning. Their lexical knowledge appears not as good as their grammatical knowledge, so that they need more strategy use to facilitate their lexical knowledge access. This partially supports the notion that for L2 students, the deficiency of lexical knowledge, is chiefly responsible for their poor reading performance, as some L2 reading studies suggest (e.g., Auerbach & Paxton, 1997; Laufer & Sim, 1985; Yorio, 1971). In addition, an implication is available for English language teachers in Taiwan. Similar to what has been suggested in the last paragraph, teachers need to make more effort to aid students in how to accumulate and expand their lexical knowledge, which is addressed in Section 7.2.2.1.

6.4 Conclusion

The current study examines the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy deployment, and their multiple-choice reading comprehension test performance. With the application of structural equation modeling, this study provides the following major findings.

Firstly, the relationship among students' English language knowledge, reading and test-taking strategy deployment, and their multiple-choice reading comprehension test performance is multi-directional, and occasionally subtle and interactive. Such a relationship is attributed to (a) no single-dimensional constructs underlying English language knowledge, and reading and test-taking strategy use; (b) possible occurrences of linkages among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance.

A variation is present in the relation among students' English language knowledge, reading and test-taking strategy use, and their reading test performance across English ability levels. For the LEA group, the aforementioned relationship is slightly more complicated than that for the HEA group. More interactions among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance are observed in the LEA group than those within the HEA group.

Secondly, students' English language knowledge and strategy use contribute to their reading test performance to some extent. However, compared with that of their English language knowledge, the contribution of students' strategy employment to their reading test performance is even smaller. Further, students' use of the strategy group comprising monitoring-the-test-taking-process and taking-advantage-of-test-questions inhibits their reading test performance. Interestingly, their deployment of marking strategies neither promotes nor inhibits their reading test performance, given that it yields no effect on how well they perform the reading test. When it comes to strategy use, what matters is the appropriate employment of a strategy group encompassing several strategies with diverse functions for a given task. The construct of the effect of strategy use on a given task is multi-dimensional, not limited to strategy use of itself.

While both groups' English language knowledge and strategy use contribute to their reading test performance to a certain degree, there are several discrepancies across English ability levels. First of all, a cross-group discrepancy is present in the size of the contribution that lexical knowledge makes to students' reading test performance. Lexical knowledge in the LEA group is more beneficial to reading test performance than the case in the HEA group. Also, students' strategy use varies to some extent in the reading test across English ability levels. The HEA group show an inclination to employ more partial monitoring strategies, while the LEA group tend to invoke more test-taking and marking strategies. The LEA group's employment of the strategy group covering monitoring-the-test-taking-process and taking-advantage-of-test-questions strategies is directly detrimental to their performance on the reading test. But this is absent in the HEA group. Finally, the LEA group's employment of marking strategies promotes their reading test performance through English language knowledge, while the HEA group's does not.

Thirdly, the interactive relationship between English language knowledge and strategy use is present in a multiple-choice reading comprehension test. Intriguingly, students' English language knowledge promotes or inhibits their strategy use, but their strategy use always contributes to their English language knowledge. Although the aforementioned interactive relation is captured in the HEA and the LEA groups, how English language knowledge interacts with strategy use is not completely the same across these two groups.

Finally, a language threshold is present for students' deploying some reading and test-taking strategies to contribute to their reading test performance. However, such a language threshold differs across English ability levels. For the HEA group, this language threshold is not manifested.

With these valuable findings, in the following chapter, I will address the implications and limitations of this study.

CHAPTER SEVEN
CONCLUSION

7.1 Introduction

This study aimed to explore the relationship among students' English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance. Motivated by the problems encountered during my past teaching life in Taiwan, and inspired by implications given by a number of previous reading or test-taking strategy research and L1-L2 reading studies, I undertook this quantitative study with structural equation modeling as my data analysis methodology.

In the following sections, I first discuss the implications for the college entrance examination center in Taiwan and for English language teachers at the senior high school level in Taiwan. Next, I discuss methodological implications of the study. Finally, I explain limitations of the study and provide recommendations of further research.

7.2 Implications

On the basis of the findings in the current study, three implications are drawn. The first implication is for the college entrance examination center in Taiwan; the second one is for English language teachers at the senior high school level in Taiwan; and the third one is pertinent to methodological implications. In the next subsections, I will, at first, discuss the implication for the college entrance examination center in Taiwan.

7.2.1 Implications for the college entrance examination center in Taiwan

As outlined in Sections 6.2, the present study displays that English language knowledge (i.e., lexical knowledge and grammatical knowledge) and strategy use exert effects on multiple-choice reading comprehension test performance in multi-directional ways and with differential effect strengths. These imply a picture of how Taiwanese senior high school students generally count on English language knowledge and strategy employment in order to well perform a reading comprehension subtest of the English component at the Senior High Academic Ability Examination. This picture is outlined as follows.

During the reading test, driven by a goal, students are aware of what they are going to do first. They access their lexical knowledge as well as grammatical knowledge greatly for reading passages in English smoothly or performing the test well. Compared with lexical knowledge, they depend more on grammatical knowledge when sitting this EFL reading test. They tend to conduct the local reading and the global reading. For example, they use their words to interpret the meaning of the sentence or predict what is coming next in the reading passage. They often check their comprehension of and modify their hypothesis about what they read. When not getting a grip on the input, they probably repeat the incomprehensible parts. They also retrieve their comprehended parts of the passage or their cognitive resources and link them with what they are processing. They will conduct marking when they do not make sense of the input. In addition, they may capitalize on strategies in a combination manner so as to promote their test performance. For example, they utilize a strategy group covering monitoring, repeating, retrieving-linking and managing-the-test strategies in concert with a strategy group encompassing construct-the-meaning and evaluating strategies to contribute to their reading test performance. Moreover, they draw upon their English language knowledge to assist in their monitoring the reading and the test-taking processes, managing the test, making an evaluation and tapping into test questions, so that their reading test performance is boosted. However, their strategy deployment at times inhibits their test performance. To illustrate, their employment of the strategy group subsuming monitoring-the-test-taking process and taking-advantage-of-test-questions strategies is directly detrimental to their reading test performance. They also strive to make appropriate sense of test questions, take heed of test time, and spend more time on challenging questions.

The abovementioned picture with students' mental procedures or behaviors being shown can be taken as an indication that students make the effort to construct the meaning of the input by interacting their own cognitive resources with the passage or test questions. While students capitalize on managing-the-test and taking-advantage-of-test-questions strategies during the reading test-taking process, it still can be argued that students attempt to comprehend the reading passage, since the employment of these strategies generally suggests that students are engaged in meaning construction to a certain degree. To illustrate, the strategy of *when I read a sentence, I noticed it was related to test questions* manifests that students attempt to grasp the meanings of test questions and a sentence or sentences they are processing. Additionally, they need to construct the mental representation of test questions and a sentence or sentences encountered to some extent, so that they can employ this strategy effectively and appropriately. Thus, we can conclude that senior high school students (third-graders), when sitting this reading comprehension test, put a certain level of effort into having a grip on the passage at the local and global levels and test questions in order to obtain the main idea of the passage, look for facts or details, or draw inferences, despite sometimes their comprehension of the input including the result of their interacting with test questions or options. For Taiwanese third-graders of senior high, validity is present to some extent in this reading comprehension test. This is fair information for the college entrance examination center in Taiwan, given that the reading passages and test items involved in the multiple-choice reading comprehension test were drawn from the reading comprehension subtest of the English component at the Senior High Academic Ability Examination (SHAAE).

Furthermore, the present study suggests that Taiwanese senior high school students' English language knowledge and strategy use have an impact on their multiple-choice reading comprehension test performance through multiple paths (see Section 6.2 for details). This implies that students access their English language knowledge and tap into strategies in a complex and strategic way to tackle a multiple-choice reading comprehension test. There is a need for the college entrance examination center in Taiwan to provide a model which profiles the paths that Taiwanese senior high school students (test-takers) follow to reach a possible answer with their English language knowledge and their strategy employment in the multiple-choice reading comprehension subtest of the English component at the SHAAE. Rogers and Bateson's (1991; 1994) model of test-taking behavior of skillful test-takers, discussed in Section 3.6.2.2 and the model regarding the relation among English language knowledge, strategy use and reading test performance, provided in the current study can function as starting points. Also, Bachman and Palmer's (1996) model of language ability can be a reference point. The preliminary model that Taiwanese senior high school students reach an answer in an EFL multiple-choice reading comprehension test is suggested as follows.

Within this suggested model (see Figure 7.1), English language knowledge is concerned with information related to English and stored in students' memory for their language use. Strategy use concerns students' deployment of mental or behavioral activities that are directly or indirectly related to their test performance. Metacognitive awareness relates to students' being able to think about their test-taking processes. Information from reading passages refers to the outcome that students obtain after they process reading passages with their English language knowledge and strategy use. Information from test questions and options relates to the outcome that students gain after they process test questions and options with their English language knowledge and strategy use.

The suggested model consists of three test-taking stages. The first stage is represented by a rectangle at the top of the model in Figure 7.1. This stage is concerned with students' reflecting on how they are going to approach reading tests. The second stage is symbolized by two rectangles at the middle of the model, meaning different approaches which students adopt to deal with tests encountered. The third stage is characterized by two rectangles that denote approaches that students further utilize to arrive at possible answers.

At the first stage, English language knowledge, strategy use, multiple-choice reading comprehension tests and metacognitive awareness are included, as shown in the top rectangle in Figure 7.1. The former three components interact with one anther. This interactive relation is indicated by a bidirectional arrow. Metacognitive awareness operates in the entire test-taking process and influences the interaction of English language knowledge, strategy use and multiple-choice reading comprehension tests, which is symbolized by a broken line circle. In this phase, students undertake a preliminary and brief interaction with tests encountered with their English language knowledge and strategy use. Then, they decide how to approach tests. This comes to the second stage. (They perhaps skip this phase and enter the next stage according to their previous test-taking experience.)

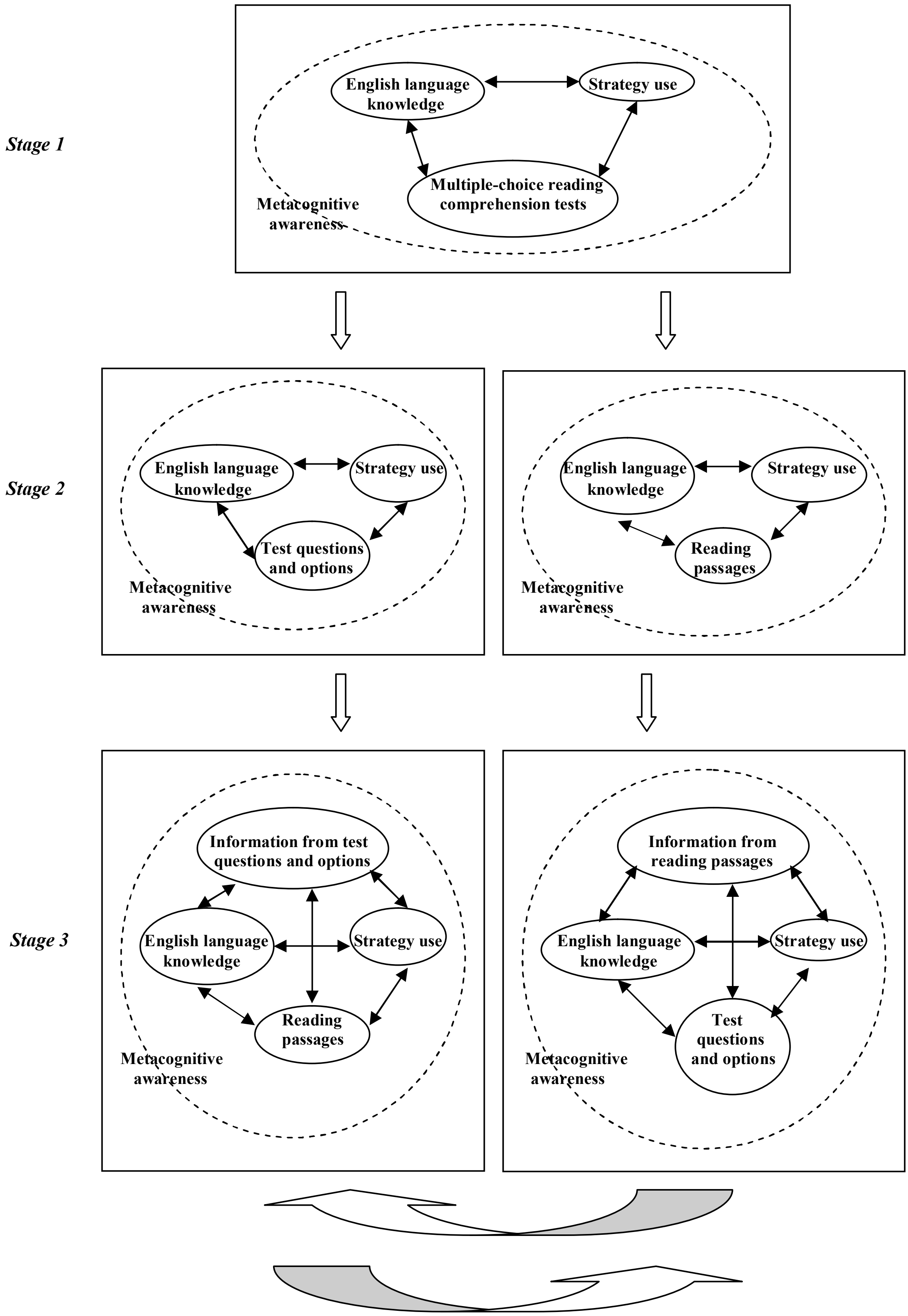


Figure 7.1 The proposed model that Taiwanese senior high school students arrive at an answer in an EFL multiple-choice reading comprehension test

At the second stage, students may decide to process reading passages at first, as indicated by the right rectangle at the middle of the model. Four components are contained in the rectangle: English language knowledge, strategy use, reading passages and metacognitive awareness. The former three components interact with one another. Such an interactive relationship is characterized by a bidirectional arrow. Metacognitive awareness still functions and impacts upon the interaction of English language knowledge, strategy use and reading passages. This rectangle suggests that students access their English language knowledge or deploy strategies if needed to get a grip on what is read. In addition, what is processed triggers students' English language knowledge access and their strategy use.

On the other hand, students probably choose to process test questions and options at first, as shown by the left rectangle at the middle of the model. Four components are covered in the rectangle. Except test questions and options which replace reading passages, other components are the same as those mentioned above. This rectangle indicates that students get access to their English language knowledge or invoke strategies if necessary to comprehend the meanings of test questions and options. Additionally, test questions and options trigger students' English language knowledge access and strategy employment.

After making sense of either reading passages or test questions and options, students proceed to the third stage. If students process reading passages first, then they move to and try to answer test questions, as represented by the right rectangle at the bottom of the model. Within this rectangle, five components are available: information from reading passages, English language knowledge, strategy use, test questions and options, as well as metacognitive awareness. The four components interact with one another. This rectangle suggests that students attempt to answer test questions or get some clues from test questions and options with their English language knowledge, strategy use and information they gain from reading passages. Furthermore, information revealed from test questions and options is integrated into students' mental representation of the passage, which has been constructed, and also influences their further understanding of the passage, English language knowledge access, and strategy deployment.

By contrast, if students process test questions and options first, then they shift to reading passages to make sense of passages, as characterized by the left rectangle at the bottom of the model. There are also five components in this rectangle. Except information from test questions and options which replace information from reading passages, other four components are the same as those stated above. This rectangle indicates that students make an effort to grasp reading passages with their English language knowledge, strategy deployment and information emerging from test questions and options. Moreover, what is constructed about the mental representation of the passage is integrated with information obtained from test questions and options, and also impacts upon students' finding answers, accessing English language knowledge and employing strategies.

Finally, during a test-taking process, students may at first read part of reading passages and then move to test questions to see whether they can answer test questions or get some clues from test questions and options to promote their understanding of what they read. Then, they move back to reading passages to process them further. They repeat the passage-and-question/option procedure. Or they may first read test questions and options, then shifting to reading passages to comprehend part of passages in order to answer test questions. Then, they proceed to test questions again to answer them or try to obtain more information from test questions and options. They repeat the question/option-and-passage procedure. These two procedures are indicated by two arrows at the bottom of the model ( and ).

Although not addressing how students intact with reading passages, test questions and options to obtain their mental representations of the input in detail and how students interact with different types of test questions, this suggested model provides a preliminary frame of reference for the college entrance examination center in Taiwan. It facilitates their producing a mature model that describes how Taiwanese senior high school students reach answers in a multiple-choice reading comprehension test. As Rupp *et al.* (2006) call for, theoretical models which profile L2 test-takers' test-taking process need to be constructed in order to provide better insights into how L2 test-takers arrive at answers in a reading test situation. With such models, the college entrance examination center in Taiwan can have a clear understanding of what their multiple-choice reading comprehension tests measure. This understanding is conducive to enhancing the validity of multiple-choice reading comprehension tests they develop. In addition, language teachers, educational decision-makers and students (test-takers) themselves can be adequately informed of test results which usually play a crucial role in students' (test-takers') future.

However, one model is not enough. We have learned from the present study that the HEA group vary from the LEA group to a certain degree in the extent to which they rest on lexical knowledge and grammatical knowledge and in their strategy employment within this reading test-taking context (see Sections 6.3.1.1, 6.3.1.2 and 6.3.2). How possible answers are reached in a multiple-choice reading comprehension test is not completely the same across English ability levels. It is necessary, for the college entrance examination center in Taiwan, to formulate different models for groups with high English ability and those with low English ability to respectively characterize the way that they reach plausible answers with their cognitive resources. After all, the one-size-fit-all model, as Purpura (1999) suggests, does not represent appropriately how groups with divergent English ability tackle a given test with their cognitive resources.

Finally, this study shows us that students' strategy deployment has an effect on their performance on the multiple-choice reading comprehension test (see Table 4.4, p. 118). As noted earlier, this reading test subsumes reading passages and test items drawn from the reading comprehension subtest of the English component at the Senior High Academic Ability Examination (SHAAE). Then, the college entrance examination center in Taiwan, in maximizing the fairness of the test, should list and publicize strategies deployed by students in the reading comprehension subtest of the English component at the SHAAE. By doing so, students who are prospective test-takers (usually third-graders of senior high schools) for the SHAAE can make reference to these strategies. The possibility, then, can be minimized that students who are equipped with knowledge of reading and test-taking strategies have an inequitable advantage over those who are not, similar to what Ellis and Ryan (2003) suggest. Although strategy deployment varies from individual to individual, students who are prospective test-takers for the SHAAE are entitled to be informed of the presence of strategies that can be taken advantage of in a reading comprehension test in a multiple-choice format, given the fairness of the test.

7.2.2 Implications for English language teachers at the senior high school level in Taiwan

The present study aims to provide a clear picture of the relationship among students' English language knowledge, strategy use, and their reading test performance. While the context where data is collected is limited to southern Taiwan, the findings shown in the current study provide two implications for English language teachers at the senior high school level in Taiwan, which will be addressed in the following.

7.2.2.1 A flexible adjustment of the order that English new words and reading texts are taught

As we have learned from the current study, students' grammatical knowledge has more positive impacts on their multiple-choice reading comprehension test performance than lexical knowledge (see Table 4.4, p. 118). Stated another way, students' grammatical knowledge contributes more to their reading test performance than lexical knowledge. This implies that in this multiple-choice reading comprehension test, Taiwanese senior high school students do not possess an appropriate amount of lexical knowledge for them to count on, in comparison with their grammatical knowledge. Several L2 reading studies (e.g., Auerbach & Paxton, 1997; Jiménez, *et al*, 1996; Yorio, 1971) have suggested that L2 students' deficiency of L2 lexical knowledge is the main obstacle to their smooth L2 reading. Similarly, Taiwanese senior high school students' lacking sufficient lexical knowledge is their weakness in performing the multiple-choice reading comprehension test with reading passages and test items drawn from the reading comprehension subtest of the English component at the SHAAE. Yorio's (1971) study has shown that vocabulary, compared with grammar, is a more challenging part for L2 learners in their L2 reading. Yorio explains that L2 learners can acquire most grammatical knowledge of an L2 and even master it through persistent learning because grammatical knowledge is more systematic and finite. However, it is quite difficult for L2 learners to master lexical knowledge due to its less systematic and infinite nature. In English language classes, Taiwanese teachers, in addition to maintaining a certain amount of focus on developing students' grammatical knowledge, need to put more effort into assisting students in how to expand and consolidate their lexical knowledge.

Without detailing how to improve students' lexical knowledge (see Folse, 2004; Hunt & Beglar, 2005 for more information), vocabulary learning strategies, as Fan (2003) argues, is useful in cumulating and expanding students' lexical knowledge. As listed in literature pertinent to vocabulary learning strategies (e.g., Fan, 2003; Gu & Johnson, 1996; Hunt & Beglar, 2005; Schmitt, 1997; 2000), vocabulary learning strategies consist of learning new words with the context provided, conducting extensive reading, familiarizing oneself with prefixes or suffixes, guessing the meaning of an unknown word from context and so forth. Aside from helping students accumulate and consolidate their lexical knowledge, through learning and employing these strategies, students' vocabulary inferencing ability is also fostered, as several L2 researchers suggest (e.g., Fraser, 1999; Hunt & Beglar, 2005; Kern, 1989). In other words, students, by their own cognitive resources, are more able to infer the meanings of unfamiliar words with the context with which the words are embedded. Taiwanese English language teachers can develop and advance students' vocabulary inferencing ability in regular English language classes with a flexible adjustment of the order that they teach reading texts and English new words.

According to my observation, in Taiwan, most English language teachers conduct the bottom-up approach to teach an English lesson. More specifically, they get accustomed to teaching students English new words involved in a reading text first, then proceeding to teaching the content of the reading text. Some even instruct students in sentence patterns included in a reading text first. Such an approach allows students to familiarize themselves with English new words or sentence patterns that they will encounter when reading the text. Students probably have less difficulty in comprehending the reading text being about to be taught. However, it provides few opportunities for students to develop their top-down reading ability, such as predicting, or hypothesis-making and testing that several L2 reading studies have identified (e.g., Davis & Bistodeau, 1993; Jiménez, *et al.*, 1996; Pritchard, 1990; Upton & Lee-Thompson, 2001; Yang, 2006). Then, after three years, students do not get used to or have enough confidence in working out the meanings of unfamiliar words from the surrounding context. Their vocabulary inferencing ability is limited. When encountering unfamiliar words during the English reading, they are not capable of deciding whether these words are keys to their comprehension and are less likely to succeed in inferring the meanings of unfamiliar words from sentences nearby[12].

English language teachers at the senior high school level in Taiwan should adjust the order that they teach reading texts and English news words, based on the cognitive difficulty of contents of reading texts. Similar to what has been adopted in some strategy instruction studies (e.g., Macaro & Erler, 2008), teachers can invite two or three students with average English proficiency to preview a reading text of an English lesson being about to be taught. If students consider the content of the reading text are less cognitive challenging, then they can conduct the top-down approach to teach the English lesson. To put it another way, they teach a reading text prior to teaching English new words or sentence patterns covered by the text.

Students can skim the entire reading text first and discuss with other students what

[12] Whether the meanings of unfamiliar words can be successfully inferred from context is subject to factors internal to students, such as English language knowledge, and factors external to students, such as the difficulty of the content of a reading text (see Hunt & Beglar, 2005, for the detailed procedures about how to enhance inference-making).

they make sense of. Teachers conclude students' discussion by providing a general idea of what the entire reading text pertains to. Then, teachers lead students to process a reading text from paragraph to paragraph by informing students of what each paragraph is concerned with, previous to their focusing on it. When encountering a new word, teachers invite students to infer its meaning with clues emerging from the immediate context or knowledge about prefixes or suffixes of English words. Such clues may be cause and effect relations or the grammatical categorization of a new word in the sentence. Students make a hypothesis about the meaning of a new word first. They then continue their reading to test whether their hypothesis is confirmed or rejected with teachers' guidance. Teachers need to offer help aptly when the clues are limited or another new word appears in the hypothesis-testing process.

With a flexible adjustment of the order that English new words and reading texts are taught, Taiwanese teachers at the senior high school level can develop students' vocabulary inferencing ability to some extent in English language classes. During the vocabulary inferencing process, students might learn that sometimes the meaning of vocabulary varies with the context in which it is involved. This is conducive to students' elaborating their lexical knowledge and further their reading comprehension – the commonly-called "depth of lexical knowledge is a necessary component of reading comprehension" (Hunt & Beglar, 2005: 33). With the increase of students' English language knowledge, senior high English teachers in Taiwan can place more emphasis on the top-down approach to teach an English lesson.

7.2.2.2 Strategy instruction and metacognitive awareness

This study has indicated that students' strategy use yields either a weak, positive or a trivial, negative effect on multiple-choice reading comprehension test performance (see Table 4.4, p. 118). That is, their strategy employment is limitedly conducive to and not always beneficial to their reading test performance. Taiwanese senior high school students' strategy use is not appropriate and effective to a certain degree in the L2 reading test-taking context. According to a number of strategy instruction studies (e.g., Amer, 1993; Auerbach & Paxton, 1997; Barnett, 1988; Carrell, *et al.*, 1989; Farrell, 2001; Kern, 1989; Macaro & Erler 2008; Nakatani, 2005; Ritter & Idol-Maestas, 1986), strategies can be instructed, and students can improve their strategy use and enhance their performance on a given task through strategy instruction. Consequently, senior high English teachers in Taiwan can implement strategy instruction in English language classes whenever appropriate to better students' strategy employment and further promote their performance on reading tests.

Some reading strategy research has suggested that metacognitive awareness carries weight in students' strategy deployment during the reading process (e.g., Baker & Brown, 1984; Carrell, 1989; Jiménez, *et al.*, 1996; Mokhtari & Reichard, 2004; Sheorey & Mokhtari, 2001; Yang, 2006). Such awareness allows students to reflect on their reading process and strategy employment, so that their strategy use is able to effective and conducive to reading performance. Similarly, in addition to students' need for strategy instruction, the present study indicates that metacognitive awareness is involved and influential in students' reading test-taking process (see pp. 171-172 for details). Senior high English teachers in Taiwan need to boost students' metacognitive awareness on the one hand and their strategy deployment on the other hand. Then, how do they advance students' metacognitive awareness and strategy employment at the same time?

Strategy training studies have demonstrated that effective strategy instruction enhances not only students' language performance but also their metacognitive awareness (e.g., Auerbach & Paxton, 1997; Carrell, *et al.*, 1989; Macaro & Erler 2008; Nakatani, 2005). It follows that the effort should be made to maximize the effectiveness of strategy instruction. Given little possibility of sparing extra classes for strategy instruction, Taiwanese senior high English teachers can incorporate strategy instruction into regular English language classes.

Without launching into detailing strategy training procedures (for more information see Auerbach & Paxton, 1997; Grenfell & Harris, 1999; Janzen & Stoller, 1998; Macaro & Erler 2008; Winograd & Hare, 1988), at first, senior high English teachers in Taiwan can have students report what strategies they invoke for performing a reading task, as suggested in strategy-related studies (e.g., Carrell, 1998; Purpura, 1999). More specifically, students voice how they make sense of the incoming input, solve comprehension breakdowns and arrive at possible answers in multiple-choice reading tests. Several techniques are available to enable students to report their strategy use, such as think-aloud protocols, learning logs, interviews or questionnaires. Among these methods, a questionnaire is an efficient means to allow a large number of students to reflect on their strategy use at a time. In his study, Purpura (1999) suggests that L2 learners can assess their strategy use by filling in his validated cognitive and metacognitive strategy use questionnaires. Similarly, given the number of students in a class[13] and heavy teaching loads, senior high English teachers in Taiwan can administer a validated questionnaire used in the current study to students to fill in. By doing so, students can have a general understanding of what strategies they currently make use of when sitting a multiple-choice reading comprehension test. That is, students' metacognitive awareness of their strategy use is activated.

In order to enhance the effectiveness of strategy instruction, Taiwanese senior high English teachers should also draw students' attention to the utility of strategy use. Teachers' explanation and modeling are means to serve this purpose (Janzen & Stoller, 1998; Winograd & Hare, 1988). However, such an approach can invite some frustrations for teachers. Farrell (2001) reported a study in which a teacher attempted to instruct high school students in new reading strategies. However, he felted frustrated at his initial stage of strategy instruction because of students' lukewarm response. This is attributable to the fact that students themselves possibly were not really aware of the benefit they would obtain from the deployment of the strategies being taught, even although the teacher had described the usefulness of the strategy use. Students could feel that this top-down (teacher-dominated) approach of strategy instruction made no difference to regular language instruction.

A bottom-up approach (student-dominated) of strategy instruction functions as an alternative way. More specifically, students themselves experience strategy utilization and appreciate the utility of strategy use by performing a reading task with information about strategies which they can tap into to grapple with the reading task. Under this approach, students develop, evaluate and modify their strategy use with appropriate assistance and regular feedback from teachers or peers. Teachers function as a facilitator and coordinator. This approach is individual oriented, as just strategy use is subject to individuals. Students self-adjust their strategy employment learning with their current knowledge or ability pertaining to strategy use. Given that individual variations in strategy deployment are taken into account, there is a high likelihood that students are more aware of the value of strategy use, which is further conducive to strategy training being productive.

As we have learned from the current study, Taiwanese senior high school students are strategic readers/test-takers to some extent (see Section 6.2.1.2 for details). Based on this point and what is mentioned above, rather than teachers' demonstrating how to deploy strategies to process reading passages, as suggested in strategy training studies

[13] Usually, there are appropriately forty or more students in a class. Each English teacher usually teaches two to four classes.

(e.g., Janzen & Stoller, 1998; Macaro & Erler 2008; Winograd & Hare, 1988), students take a simulated multiple-choice reading test to go through strategy deployment, then realize the usefulness of strategy use and further improve their strategy use. Senior high English teachers in Taiwan can cooperate with each other and prepare simulated multiple-choice reading comprehension tests. Reading passages and test items can be drawn from the previous reading comprehension subtest of the English component at the Senior High Academic Ability Examination. Information on how to approach and deal with this type of test is provided with reading passages and test items. Such information consists of local and global reading strategies, managing-the-test strategies, monitoring strategies and so on. These strategies are presented in a way that they are categorized into several groups. Each group including strategies with different functions serves discrepant purposes. For example, a utilizing-test-questions group may cover comprehension-checking, evaluating, retrieving-linking and memorizing strategies. With the information given, students sitting simulated multiple-choice reading comprehension tests are more likely to make sense of the input and reach answers to test questions successfully. This leads students to gain a greater understanding of the fact that they can tackle multiple-choice reading comprehension tests in such an appropriate and strategic way that their reading test performance can be promoted. The understanding motivates them to be more aware of their reading and test-taking process and thereby to make appropriate use of strategies.

In addition, senior high English teachers in Taiwan can have students discuss with others their strategy employment in a simulated reading test. Students' sharing their successful or unsuccessful experiences in strategy deployment during the reading test with each other develops students' metacognitive awareness of their strategy use and benefits all students with different English ability, as is the case in Auerbach and Paxton's (1997) study. Through this sharing, students with high English ability can have a better understanding of their strategy use, and in turn employ their strategies in a more effective and skillful way. On the other hand, students with low English ability can reflect on their own strategy deployment, get a clear picture of their drawbacks to strategy use and expand their strategy repertoires by means of listening to others' successful experiences in deploying some strategies.

During the discussion, students may convey that even though they deploy strategies, sometimes their incomprehension is still present and such incomprehension obstructs them from arriving at answers to test questions. On such an occasion, senior high English teachers in Taiwan might remind students the following points. Firstly, strategy use, while beneficial to their reading comprehension test performance, is not an elixir, given that its contributions to reading test performance may not be as many as expected and is susceptible to factors internal to users (e.g., their L2 proficiency) or external to users (e.g., tasks given). Furthermore, possessing certain amounts of English language knowledge is a prerequisite since it facilitates their processing texts smoothly, employing some strategies appropriately and performing reading tests well (Anderson, 1991; Purpura, 1997; 1999; Rogers & Bateson, 1991; 1994; Yang, 2002; 2006). Reading and test-taking strategies, after all, function as a possibility to help them do away with obstacles to their reading comprehension and boost their performance on reading tests. Then, students will get a deeper understanding of the fact that English language knowledge is still crucial, even though they receive strategy instruction and deploy strategies in their English reading.

After class, students are encouraged to write learning logs to document what they have learned about their strategy use and how they feel about what they learned, as implemented in some strategy instruction research (e.g., Auerbach & Paxton, 1997; Shen, 2004). By doing so, students are more metacognitively aware of their strategy learning and strategy instruction can be more effective. If possible, English teachers check students' learning logs, especially the logs of those with poor reading performance, given that they are, as Kern's (1989) work has implied, the ones who need strategy instruction most.

As for when to initiate strategy instruction, a suggestion is offered that senior high English teachers in Taiwan commence strategy instruction when students are in their second grade of senior high. In students' first year of senior high, teachers can assist in students' accumulating their English language knowledge (e.g., lexical knowledge and grammatical knowledge) and instill into students the concepts of how to read properly. Given teaching loads and the number of students in a class, teachers can implement strategy training once every three or four weeks. Furthermore, seeing that several L2 researchers have pointed out that it takes a certain period of time to enhance students' strategy use (e.g., Auerbach & Paxton, 1997; Farrell, 2001; Janzen & Stoller, 1998), teachers should implement strategy instruction until students finish their senior high. Then, strategy instruction can be effective and productive. What students learn about strategy deployment is very likely to be transferred and conducive to their future learning in the university where students need to read textbooks or journals in English to some extent.

What is suggested about strategy instruction is not a fixed approach being applied to different senior high schools in Taiwan. English teachers can make an adjustment based on their real teaching context. We expect that the strategy instruction guidelines mentioned here can provide directions for senior high English teachers in Taiwan. With such directions, something different is added in English language classes to promote students' reading performance. Prior to proceeding to the next section, caveats should be given that as strategy instruction is incorporated into language classes, "students and teachers alike should maintain a clear focus on the final learning goal, and not pursue strategy training in and of itself" (Purpura, 1999: 186). After all, "strategy use is a means to the broader goal of acquisition or performance just as language learning is a means to the broader goal of communication" (ibid.).

7.2.3 Methodological implications

Reviewing the existing literature germane to L2 reading and language assessment, I found that both qualitative and quantitative studies had been conducted to investigate strategy use in L2 reading or L2 reading tests, the linkage between strategy use and language knowledge, or the relation among language knowledge, strategy use and performance on L2 reading tasks/tests. Qualitative studies manifest valuable information on how readers/test-takers approach a given L2 reading task/test and deploy strategies to resolve their incomprehensible parts in their L2 reading or to optimize their reading test performance (e.g., Cohen, 1984; Cohen & Upton, 2006; 2007; Dollerup *et al*., 1982; Hosenfeld, 1977; Nikolov, 2006; Rupp *et al*., 2006). With regard to quantitative studies, regression analysis has been performed to show that L2 proficiency/L2 language knowledge or strategy use contributes to L2 reading test performance (e.g., Anderson, 1991; Bossers, 1991; Carrell, 1989; Kobayashi, 2002; Padron & Waxman, 1988; Usó-Juan, 2006). Correlation analysis has also been applied to investigate the relationship between strategy use and reading task/test performance in the L2 setting (e.g., Barnett, 1988; Phakiti, 2003). In addition, t-tests, analysis of variance or multivariate analysis of variance have been carried out to explore strategy use variations across groups with different L2 reading task/test performance or conditions in which tasks with discrepant difficulty levels are used (e.g., Oxford *et al*., 2004; Phakiti, 2003).

While substantive information was offered, I noticed that such studies failed to exhibit a picture of how readers'/test-takers' English language knowledge, strategy use and their reading test performance interacted with one another in a single modeling framework. More specifically, they could not present effect or correlational paths among variables of interest in a model. Consequently, I was searching for a methodology which was able to manifest the linkage of variables of interest in a model.

Perusing the related literature, I found that a structural equation modeling (SEM) approach had been applied to investigate several issues in the field of L2 learning. For example, Sasaki (1993) adopted SEM to study the relationships among L2 proficiency, foreign language aptitude, and intelligence. With an SEM approach, Kunnan (1995) investigated the effects that test-taker background characteristics exerted on L2 test performance. Purpura (1997, 1998b, 1999) examined the relation between strategy use and L2 test performance by using SEM. Schoonen *et al*. (1998) explored the relationship between primary school students' metacognitive and language-specific knowledge in their native and foreign language reading test performance, with the application of SEM. Van Gelderen *et al*.'s (2004) performed SEM to examine the connection amongst linguistic knowledge, processing speed and metacognitive knowledge in L1 and L2 reading. With an SEM approach, In'nami (2006) explored the relationship between test-takers' test anxiety and their L2 listening test performance. Finally, Shiotsu and Weir (2007) conducted SEM to investigate to what extent syntactic knowledge and lexical knowledge contributed to L2 reading test performance. From these studies, I found that SEM allowed the relations between observed variables and latent variables and those among latent variables to be inspected and shown in a single modeling framework. It was this multivariate analytic procedure that was appropriate for my study.

Within the current study, from the perspective of a structural equation modeling (SEM) approach, I investigated the relationship among Taiwanese senior high school students' English language knowledge, strategy use, and their reading comprehension test performance. In order to examine the aforementioned relation, I first conducted exploratory factor analysis to validate measuring instruments administered and to obtain their underlying components (i.e. constructs). During the process, sometimes I felt it difficult to decide the number of and label components. Upon reflection, I learned that it would have been better to decide categorically the types of components which measuring instruments were intended to assess on theoretical underpinnings or related studies, prior to their being developed. Two components were extracted from a 56-item English language knowledge test (see p. 250 for details), four components from a 72-item strategy use questionnaire (see p. 255 for details), and two components from a 17-item reading comprehension test (see p. 263 for details). With these results, I was aware that exploratory factor analysis was instrumental in summarizing the substantial amount of the data collected from measures to acquire the adequate number of components to represent the data.

Next, I performed confirmatory factor analysis to test the relationship between observed variables and latent variables within the measurement model of English language knowledge and that of reading and test-taking strategy use. With the results, I found that confirmatory factor analysis conducted by SEM was useful for giving insights into the extent to which observed variables (i.e., test items or questionnaire items) could represent latent variables (i.e., components extracted from exploratory factor analysis). Additionally, a close examination of the measurement model of reading and test-taking strategy use indicated that there were error-correlations manifested in the model. This led to my realization that SEM was powerful enough to capture and present measurement errors in a single modeling framework to provide information on how well measures worked, given that measurement errors were taken into parameter estimation in the SEM analysis.

After the measurement models of English language knowledge, strategy use and reading test performance were formulated, I constructed the full latent variable model regarding the relationship among students' English language knowledge, strategy use, and their reading test performance with the application of SEM. In the model generating course, a model with poor goodness-of-fit was first yielded, predicated on theoretical underpinnings and related studies. This led me to attempt to adjust the model frequently in order to obtain a model with satisfactory goodness-of-fit. However, upon reflecting on caveats given by some SEM researchers (e.g., Chiu, 2006; Hung, 2002), I made as few adjustments as possible for the hypothesized model for fear that excessive manipulation from a researcher is involved. Within the accepted model, I noticed that not only latent variables but also observed variables exerted an effect on latent variables. This demonstrates that SEM is so powerful that it is capable of capturing the possible relationships among variables encompassed in a postulated model.

Rather than just taking the finding produced from a single group model, then applying it to groups with discrepant English ability, I conducted the separate group analysis by formulating the model regarding the relation among English language knowledge, strategy use, and reading test performance for groups with different English ability. With cross-group commonalities and variations in the aforementioned relation being located, I realized that SEM was useful for constructing different models for groups with discrepant attributes in order to see whether differences were present across groups. Thereby, I would have an understanding of whether group attributes made a difference to a hypothesized relationship among variables of interest.

Finally, I carried out the simultaneous group analysis to test whether parameters on the paths shared by both groups were equivalent. Among studies related to strategy use, only Purpura's (1998b, 1999) conducted the simultaneous multi-sample analysis. I found that there were no statistical differences in the magnitudes of some effects for the paths shared by both groups, despite the presence of the apparent cross-group discrepancy in the magnitudes of the effects, which was pinpointed in the separate group analysis. For instance, the path that grammatical knowledge had a direct effect on multiple-choice reading comprehension test performance was shared by groups with different English ability in the separate group analysis. The effect of grammatical knowledge on multiple-choice reading comprehension test performance in the HEA group was stronger than that in the LEA group (.459 vs. .315, see Table 5.2, p. 135). However, this cross-group difference was rejected in the simultaneous group analysis. Without the simultaneous group analysis, I would have taken this cross-group variation as a final cross-group discrepancy. In this respect, SEM indicates its robust and useful ability to test cross-group differences in the relation among variables investigated from a stringent statistical perspective by performing a cross-group invariance test with equality constraints imposed on parameters of interest. In so doing, more robust evidence is provided.

To conclude, with substantive findings offered by the current study (see Chapter Six for details), the utility of multivariate analytic procedures for examining the relationship among students' L2 language knowledge, strategy employment and their reading test performance is demonstrated. Despite some limitations of SEM (see the fifth and the sixth points mentioned in Section 7.3), SEM still functions well as a methodology to study the relation among variables under investigation. Within an SEM model, a lucid picture of how a variable has a direct or indirect effect on another can be shown. Further, SEM is also able to show information on whether a postulated relation among a set of variables is invariant across groups with different attributes. This is important, since it can display cross-group discrepancies and provide more empirical evidence for the hypothesized relation, based on the data gathered from another group of participants. In the future, I will still carry out this methodology when investigating the relationship among a large group of variables.

7.3 Limitations of the study

Although the present study has produced substantive findings, its design is not without flaws. First of all, the findings here are generated from senior high school students in the EFL context. These participants learned English as a foreign language and had learned it at least for five years at the time of the study. They were all 17 years of age or older. As a consequence, these findings may not be generalized to other contexts, such as the ESL (English as a second language) setting or other populations, such as English language learners at the elementary school level. Additionally, based on my personal contacts and availability of participants, participants in the current study were selected only from six senior high schools in the south region of Taiwan; thus, the generalizations of the findings to similar populations should be treated circumspectly.

Secondly, the findings here simply throw partial light on the relationship among students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance, given the limitation of data collection instruments. To illustrate, in the present study, reading test performance was measured by a multiple-choice reading comprehension test with 17 test items; therefore, the findings here can only extend to this type of reading test. Moreover, English language knowledge was limited to lexical knowledge and grammatical knowledge which respectively gauged by vocabulary and grammar subtests. Accordingly, the findings should be treated with caution when English language knowledge is referred to. Finally, while the strategy item pool was developed based on the results of retrospective interviews conducted on participants similar to those in the present study and strategy items identified or used in other strategy-related studies (see p. 71 for these studies), what the strategy use questionnaire covered was limited. As a result, the collected data from the strategy use questionnaire fails to provide a complete picture of EFL students' strategy employment in this test-taking context or other contexts.

In addition, information on participants' attributes of interest is obtained indirectly by administering measures to participants. Neither the reading comprehension test nor the English language knowledge test can completely mirror participants' reading ability or language knowledge. Also, a self-report questionnaire can not reflect their mental activities comprehensively in the reading test-taking setting. Accordingly, what the reading test, the English language knowledge test, and the questionnaire capture is part of participants' reading test performance, English language knowledge and strategy deployment. Given this, it is acknowledged that internal validity might be diminished.

Based on what is mentioned here and in the previous paragraph, the findings of this study simply manifest the partial relationship among students' English language knowledge, reading and test-taking strategy use, and their multiple-choice reading comprehension test performance.

Thirdly, in the present study, participants' English ability was determined by the scores of their self-rating English ability, rather than by a standardized measure. In spite of many attempts made to justify this decision, great caution should be taken, when the findings regarding commonalities and differences in the relationship among English language knowledge, strategy use, and reading test performance across English ability levels are referred to.

Fourthly, the current study categorized reading and test-taking strategies into four groups through the application of exploratory factor analysis. Doing so provides an insight into what components underlie the strategy use data. It also makes it possible that a measurement model of reading and test-taking strategy use can be further constructed by SEM. However, this approach presents a flaw. That is, sometimes it is difficult to label a group precisely, given that the group subsumes more than one salient feature[14]. For example, the *monitoring, directing attention and managing the test* process consists of three salient features: monitoring, repeating and managing-the-test. It leads to the fact that the SEM results about the relationship among students' English language knowledge, strategy use, and their reading test performance sometimes can not be explained precisely despite an effort being made. Furthermore, post-hoc naming results in the fact that the four strategy use processes[15] in the present study cannot categorically correspond to reading processes in Pressley and Afflerbach's (1995) model of constructively responsive reading and test-taking processes in Rogers and Bateson's model of expert test-takers' test-taking behavior (1991; 1994)[16]. It follows that construct validity of a strategy use

[14] I tried to solve this problem in two ways. Firstly, some strategies were removed from a group to another group, so that it could be easier to label a group. However, such approach led to the fact that the results of exploratory factor analysis could not be held, as the results were further tested by confirmatory factor analysis with the use of SEM. The second way was to drop some strategies. However, doing so reduced the number of strategies in a strategy group and so did the reliability of the strategy group. Further, strategies in a strategy group were further separated into two or more strategy subgroups for the SEM analysis. Reducing the number of strategies in a strategy group also impinged upon the reliability of a strategy subgroup. Consequently, I finally did not adopt these two means.

[15] These four processes consist of (a) the *monitoring, directing attention and managing the test* (MDAMT) process; (b) the *constructing the meaning and evaluating* (CME) process; (c) the *monitoring and utilizing test questions* (MUTQ) process; (d) the *evaluating and marking* (EM) process.

[16] A strategy use questionnaire was predicated on Pressley and Afflerbach's (1995) model of constructively responsive reading and Rogers and Bateson's model of expert test-takers' test-taking behavior (1991; 1994).

questionnaire is compromised to some extent and so is that of the current study. Accordingly, the findings here should be treated with caution.

Fifthly, as noted in the last paragraph of Section 4.3.3, SEM was utilized in an exploratory manner in five aspects. Additionally, some effects captured in the current study were trivial such as those that strategy use has on reading test performance, English language knowledge yields on strategy use and strategy use exerts on English language knowledge. When these effects were referred to and interpreted, caution ought to be taken. More work with different groups of participants is necessary to be carried out in order to confirm some findings produced in the present study.

Sixthly, the findings in this study are predicated on an accepted model produced by structural equation modeling (SEM). However, there are still maybe a number of alternative models which may fit the sample data better. As a consequence, the findings here are tentative. Further, although in the current study the accepted model depicts the collected data fairly, we have no idea of whether this model is exactly true, given the limitations of SEM. All we know is that the model is accepted according to a set of the model fit indices adopted. Hence, the accepted model here is just an approximation of the true model. Caution should be taken in an attempt to discuss the findings produced from this approximation model.

Seventhly, the current study manifests causal effects between variables with the application of structural equation modeling (SEM). This multivariate analytic procedure is able to show causal effect paths in a single modeling framework – the paths can not be manifested clearly in qualitative studies. It also does not require researchers to control variables circumspectly, as researchers in experimental studies do – sometimes it is difficult to control variables precisely[17]. However, due to this, causal effects yielded from SEM should be interpreted in a conservative way. The effects observed in the current study can merely be regarded as the effects approximating those identified in an experimental study.

Finally, as noted in the last paragraph on Section 3.2, the present study is a cross-sectional survey. Information of interest was simply collected once. Clearly, the gathered

---

These four processes, generally, are compatible with reading and test-taking processes involved in the aforementioned models.

[17] For example, in an experimental study, if a researcher is interested in whether strategy use yields an effect on reading test performance, he/she usually needs to divide participants into two groups and carefully control other factors than strategy use that affect reading test performance. However, in fact, it is impossible for him/her to control all the other factors that impact upon reading test performance, given the number of these factors involved and the possibility of unknown factors included.

information is limited. Thus, the findings produced in this one-shot study should be treated carefully, when it is referred to.

7.4 Suggestions for further research

As mentioned above, the current study has some limitations; nonetheless, it provides a basis for further research to investigate the relationship between EFL students' (test-takers') characteristics and reading test performance. Several research works which may be carried out are listed as follows.

As for operationalization of latent variables, reading comprehension test performance measured only by the multiple-choice reading comprehension test, no doubt, gives a narrow view of reading test performance. Future studies can cover different types of reading tests, such as cloze tests, to operationalize reading comprehension test performance. Further, English language knowledge in this study was limited to lexical knowledge and grammatical knowledge. More types of language knowledge, such as textual knowledge, can be included in future research to operationalize language knowledge. Finally, in order to provide a comprehensive picture of EFL students' strategy use in the reading test course, more strategies can be involved in the strategy use scale in further studies.

With regard to the criterion assumed to decide participants' English ability, in the current study, participants' English ability was not assessed by a standardized measure given the considerable number of participants, limited data collection time and resources available. In a future study, participants' English ability can be gauged by a standardized measuring instrument. Then, the obtained findings can be compared with those yielded here to see whether there is any difference.

Due to the limited number of female participants, the present study did not take gender differences into account in the full latent variable model. Further studies can address whether gender differences have an influence on strategy use with the use of SEM. In addition, the simultaneous group analysis can be performed to see whether the relationship among English language knowledge, strategy use, and reading test performance are equivalent across the male student group and the female student group.

The current study only investigated the relationship among students' English language knowledge, strategy use, and their multiple-choice reading comprehension test performance from the perspective of SEM. A qualitative study with think-aloud procedures to collect the data regarding how students at the senior high school level approach a reading test can be carried out to cross-validate the findings produced in this study.

In addition, the future research can be undertaken in which more participants are recruited and their English ability is measured by an English language test (e.g. IELTS – International English Language Testing System). Then, participants are divided into several groups (e.g., three) according to the result of the test. The relationship among English language knowledge, strategy use, and reading test performance is modeled respectively for each group. A comparison is made amongst group models to pinpoint which IELTS level corresponds to the upper and the lower language thresholds for some strategy deployment with English language knowledge as a mediator to contribute to reading test performance in the EFL context.

Finally, as far as factors related to reading test performance are concerned, future research can include more other factors, such as knowledge of subject matter, test anxiety, learning attitudes, text structures or difficulty of test items to provide more insights into the relationship between reader/test-taker based factors and text/test-task based factors in the EFL reading setting. In addition, a more comprehensive picture can be got of how EFL students (readers/test-takers) with different reader/test-taker characteristics interact with a reading task/test with discrepant text/test-task attributes.

7.5 Concluding statements

The current study set out to investigate the relationship among Taiwanese senior high school students' English language knowledge, reading and test-taking strategy deployment, and their multiple-choice reading comprehension test performance with the use of multivariate analytic procedures – structural equation modeling (SEM).

With the findings discussed in Chapter Six, we draw several conclusions. First of all, Taiwanese senior high school students are strategic readers/test-takers. Their English language knowledge and strategy use contribute to their multiple-choice reading comprehension test performance. However, the contribution of their strategy use to their reading test performance is limited and even smaller than English language knowledge. Sometimes their strategy use is not appropriate and effective. There is a need for implementing strategy instruction for students to improve their strategy use in a reading test and further to boost their reading test performance. Prior to strategy instruction, they ought to accumulate their English language knowledge to some extent, so that their strategy deployment can be more appropriate and effective.

Additionally, the validity of the reading comprehension subtest of the English component at the Senior High Academic Ability Examination (SHAAE) is present to a certain degree. Also, given maximizing the fairness of the test, the college entrance examination center in Taiwan should list and publicize strategies which students can take advantage of when they sit the reading comprehension subtest of the English component at the SHAAE.

Furthermore, the present study gives more empirical evidence for Bachman's (1990) factors that influence test scores and Bachman and Palmer's (1996) model of language ability in language use and language test performance, both of which the current study is predicated on. In Bachman and Palmer's (ibid.) model, the definition of strategic competence needs to be revised.

The current study also provides more evidence for several findings, shown or suggested in previous qualitative or quantitative studies, especially in Purprua's (1997; 1998b; 1999) studies closely related to the current study. The findings consist of English language knowledge and strategy use influencing and promoting reading test performance, English language knowledge enhancing reading test performance more than strategy use, the presence of a language threshold for some strategy deployment, and so forth.

Finally, SEM is a useful multivariate, analytic procedure for investigating the relation among variables of interest in a single modeling framework. It is also helpful in examining whether the result produced in a group model is equivalent within another group model. In this way, cross-group commonalities and variations can be located and yielded results can be cross-validated.

I hope that the current study can give more insights into the relationship among Taiwanese senior high school students' English language knowledge, strategy use and their performance on a multiple-choice reading comprehension test. With such insights, different scenarios of English language teaching in Taiwan can be present.

References

Abbott, M. L. (2006). ESL reading strategies: Differences in Arabic and Mandarin speaker test performance. *Language Learning*, 56(4), 633-670.

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson, & A. H. Urquhart (Eds.), *Reading in a foreign language*. London: Longman, 1-24.

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alderson, J. C., & Urquhar, A. H. (1984). Introduction: What is reading? In J. C. Alderson, & A. H. Urquhart (Eds.), *Reading in a foreign language*. London: Longman, xv-xxviii.

Alexander, P. A., Graham, S., & Harris, K. R. (1998). A perspective on strategy research: Progress and prospects. *Educational Psychology Review*, 10(2), 129-154.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, 61(3), 315-343.

Allan, A. I. C. G. (1995). Begging the questionnaire: Instrument effect on readers' responses to a self-report checklist. *Language Testing*, 12(2), 133-153.

Amer, A. A. (1993). Teaching EFL students to use a test-taking strategy. *Language Testing*, 10(1), 71-79.

Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, 75(4), 460-472.

Anderson, N. J. (2005). L2 learning strategies. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning.* Mahwah, NJ: Lawrence Erlbaum Associates, 757-771.

Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. D. (1991). An exploratory study into the construct validity of reading comprehension text: Triangulation of data sources. *Language Testing*, 8(1), 41-66.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* vol. 2. White Plains, NY: Longman, 255-292.

Aslanian, Y. (1985). Investigating the reading problems of ESL students: An alternative. *ELT Journal*, 39(1), 20-27.

Auerbach, E. R., & Paxton, D. (1997). "It's not the English thing": Bring reading research into the ESL classroom. *TESOL Quarterly*, 31(2), 237-261.

Babbie, E. R. (2004). *The practice of social research* (10th ed.). Belmont, CA: Wadsworth.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

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

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* vol. 2. White Plains, NY: Longman, 353-394.

Barnett, M. A. (1986). Syntactic and lexical/semantic skill in a foreign language reading: Importance and interaction. *The Modern Language Journal,* 70(4), 343-349.

Barnett, M. A. (1988). Reading through context: How real and perceived strategy use affects L2 comprehension. *The Modern Language Journal*, 72(2), 150-162.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.

Bentler, P. M. (1995). EQS: Structural equations program manual. Encino, CA: Multivariate Software.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in SEM. *Sociological Method and Research*, 16(1), 78-116.

Bernhardt, E.B. (1991). *Reading development in a second language: Theoretical, empirical and classroom perspectives*. Norwood, NJ: Ablex Publishing Corp.

Bernhardt, E.B. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, 25, 133-150.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, 16(1), 15-24.

Bialystok, E. (1981). The role of conscious strategies in second language proficiency. *The Modern Language Journal*, 65(1), 24-35.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly,* 20(3), 463-494.

Block, E. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, 26(2), 319-341.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bossers, B. (1991). On thresholds, ceilings, and short circuits: The relation between L1 reading, L2 reading, and L2 knowledge. AILA Review 8, 45-60.

Bransford, J. D. (1979). *Human cognition: Learning, understanding and remembering*. Belmont, CA: Wadsworth.

Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. *The Modern Language Journal*, 89(1), 37-53.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum, 453-481.

Brown, A. L. (1987). Metacognition, executive control, self-regulation and other even more mysterious mechanisms. In F. E. Wienert, & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding*. Hillsdale, NJ: Erlbaum, 65-116.

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences*. New York: Plenum Press, 185-249.

Bügel, K., & Buunk, B. P. (1996). Sex differences in foreign language text comprehension: The role of interests and prior knowledge. *The Modern Language Journal*, 80(1), 15-31.

Byrne, B. M. (2001). *Structural equation modeling with AMOS*. Mahwah, NJ: Lawrence Erlbaum Associates.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.

Carrell, P. L. (1988). Interactive text processing: Implications for ESL/second language reading classrooms. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading*. Cambridge: Cambridge University Press, 239-277.

Carrell, P. L. (1989). Metacognitive awareness and second language reading. *The Modern Language Journal*, 73(2), 121-134.

Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics*, 12(2), 159-179.

Carrell, P. L. (1992). Awareness of text structure: Effects on recall. *Language Learning*, 42(1), 1-20.

Carrell, P. L. (1998). Can reading strategies be successfully taught? *ARAL*, 21(1), 1-20.

Carrell, P. L., & Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17(4), 553-573.

Carrell, P. L., Pharis, B. G., & Liberto, J. C. (1989). Metacognitive strategy training for ESL reading. *TESOL Quarterly*, 23(4), 647-678.

Chamot, A. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, 25, 112-130.

Chiu, H.-C. (2006). *Structural equation modeling: Theory, technology and application of LISREL*. Taipei: Shang-Yie Press.

邱皓政 (民 95)。結構方程模式: LISREL 的理論、技術與應用。台北: 雙葉出版社。

Clarke, M. A. (1979). Reading in Spanish and English: Evidence from adult ESL students. *Language Learning*, 29(1), 121-150.

Clarke, M. A. (1980). The short circuit hypothesis of ESL reading — or when language competence interferes with reading performance. *The Modern Language Journal*, 64(4), 203-209.

Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, & R. R. Jordan (Eds.), *Reading in a second language*. Rowley, MA: Newbury House, 5-12.

Cohen, A. D. (1984). On taking language tests: What students report. *Language Testing*, 1(1), 70-81.

Cohen, A. D. (1986). Mentalistic measures in reading strategy research: Some recent finding. *English for Specific Purposes*, 5(2), 131-145.

Cohen, A. D. (1998a). Strategies and processes in test taking and SLA. In L. F. Bachman, & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press, 90-111.

Cohen, A. D. (1998b). *Strategies in learning and using a second language*. London: Longman.

Cohen, A. D., Glasman, H., Rosenbaum-Cohen, P. R., Ferrara, J., & Fine, J. (1979). Reading for specialized purpose: Discourse analysis and the use of student informants. *TESOL Quarterly*, 13(4), 551-564.

Cohen, A. D., & Upton, T. A. (2006). Strategies in responding to the New TOEFL reading task (TOEFL Monograph Series Report No. 33). Princeton, NJ: Educational testing Service. Retrieved October 17, 2007, from http://www.ets.org/Media/Research/pdf/RR-06-06.phf

Cohen, A. D., & Upton, T. A. (2007). 'I want to go back to the text': Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24(2), 209-250.

Cohen, L., Manion, L., & Morrison, K. (2001). *Research methods in education* (5th ed.). London: Routledge Falmer.

College Entrance Examination Center in Taiwan. (2006a). *Ability Exams*. Retrieved May 2, 2006, from http://www.ceec.edu.tw/AbilityExam/AbilityExamPaper.htm

College Entrance Examination Center in Taiwan. (2006b). *Appointed Exams*. Retrieved July 10, 2006, from http://www.ceec.edu.tw/AppointExam/AppointExamPaper.htm

College Entrance Examination Center in Taiwan. (2006c). *Research Papers*. Retrieved July 15, 2006, from http://www.ceec.edu.tw/Research/paper_doc/ce37 /ce37.htm

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt.

Cziko, G. A. (1980). Language competence and reading strategies: A comparison of first- and second-language oral reading errors. *Language Learning*, 30(1), 101-114.

Davis, J. N., & Bistodeau, L. (1993). How do L1 and L2 reading differ? Evidence from think aloud protocols. *The Modern Language Journal*, 77(4), 457-472.

Department of Elementary Education in Taiwan. (2008). *Curricular Outline*. Retrieved May 20, 2008, from http://www.ceec.edu.tw/file/site_content/boo56/english-1.doc

Devine, J. (1984). ESL readers' internalized models of the reading process. In J. Handscombe, R. Orem, & B. Taylor (Eds.), *On TESOL '83*. Washington: TESOL, 95-108.

Devine, J. (1988). The relationship between general language competence and second language reading proficiency: Implications for teaching. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading*. Cambridge: Cambridge University Press, 260-277.

Dollerup, C., Glahn, E., & Hansen, C. R. (1982). Reading strategies and test-taking techniques in an EFL reading comprehension test: A preliminary report. *Journal of Applied Language Study*, 1(1), 93-99.

Dörnyei, Z. (2003). *Questionnaires in second language research: Construction,* *administration, and processing*. London: Lawrence Erlbaum Associates.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly*, 38(1), 78-103.

Ellis, A. P. J., & Ryan, A. M. (2003). Race and cognitive-ability test performance: The mediating effects of test performance, test-taking strategy use and self-efficacy. *Journal of Applied Social Psychology*, 33(12), 2607-2629.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

English Education Resource Center in Taiwan. (2008). *Curricular Outline*. Retrieved May 20, 2008, from http://english.tyhs.edu.tw/xoops/html/modules/dms/file_ retrieve.php?function=view&obj_id=10

Erler, L., & Finkbeiner, C. (2007). A review of reading strategies: Focus on the impact of the first language. In A. D. Cohen, & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice*. Oxford: Oxford University Press, 187-206.

Faerch, C., & Kasper, G. (1983). *Strategies in interlanguage communication*. London: Longman.

Fan, M. M. (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *The Modern Language Journal*, 87(2), 222-241.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, 27(3), 209-226.

Farrell, T. S. C. (2001). Teaching reading strategies: “It takes time!” *Reading in a foreign language*, 13(2), 631-646.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.

Folse, K. S. (2004). *Vocabulary myths*. Ann Arbor, MI: The University of Michigan Press.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21(2), 225-241.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London: Routledge.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6(1), 126-135.

Gordon, C. M., & Hanauer, D. (1995). The interaction between task and meaning construction in EFL reading comprehension tests. *TESOL Quarterly*, 29(2), 299-322.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25(3), 375-406.

Grabe, W., & Stoller, F. L. (2002). *Teaching and researching reading*. Harlow: Pearson Education.

Graesser, A. C., & Kreuz, R. J. (1993). A theory of inference generation during text comprehension. *Discourse Processes*, 16(2), 145-160.

Green, J. M., & Oxford, R. L. (1995). A closer look at learning strategies, L2 proficiency and gender. *TESOL Quarterly*, 29(2), 261-297.

Grenfell, M., & Harris, V. (1999). *Modern languages and learning strategies: In theory and practice*. London: Routledge.

Grenfell, M., & Macaro, E. (2007). Claims and critiques. In A. D. Cohen, & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice*. Oxford: Oxford University Press, 9-28.

Griffiths, C. (2003). Patterns of language learning strategy use. *System*, 31(3), 367-383.

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46(4), 643-697.

Harrison, A. (1983). *A language testing handbook*. London: Macmillan.

Hauptman, P. C. (1979). A comparison of first and second language reading strategies among English-speaking university students. *Interlanguage Studies Bulletin*, 4(2), 173-201.

Heubeck, B., & Neill, J. (2000). Confirmatory factor analysis and reliability of the mental health inventory for Australian adolescents. *Psychology Reports*, 87(2), 431-440.

Hong-Nam, K., & Leavell, A. G. (2006). Language learning strategy use of ESL students in an intensive English learning context. *System*, 34(3), 399-415.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127-160.

Hosenfeld, C. (1977). A preliminary investigation of reading strategies of successful and non-successful language learners. *System*, 5(2), 110-123.

Hosenfeld, C. (1984). Case studies of ninth grade readers. In J. C. Alderson, & A. H. Urquhart (Eds.), *Reading in a foreign language*. London: Longman, 231-244.

Hsiao, T.-Y., & Oxford, R. L. (2002). Comparing theories of language learning strategies: A confirmatory factor analysis. *The Modern Language Journal*, 86(3), 368-383.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.

Hudson, T. (1982). The effects of induced schemata on the "short circuit" in L2 reading: Non-decoding factors in L2 reading performance. *Language Learning*, 32(1), 1-31.

Hung, F.-M. (2002). *Structural equation modeling: Theory and application*. Taipei: Wu-Nan Press.

黄芳銘 (民 91)。結構方程模式: 理論與應用。台北: 五南圖書。

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading a in Foreign Language*, 17(1), 23-59.

In'nami, Y. (2006). The effects of test anxiety on listening test performance. *System*, 34(3), 317-340.

Janzen, J., & Stoller, F. L. (1998). Integrating strategic reading in L2 instruction. *Reading in a Foreign Language*, 12(1), 251-265.

Jiménez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, 31(1), 90-112.

Johnston, P. H. (1983). *Reading comprehension assessment: A cognitive basis*. Newark, DE: International Reading Association.

Jöreskog, K. G. (1993). Testing structural equation models. In K. Bollen, & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage, 294-316.

Kern, R. G. (1989). Second language reading strategy instruction: Its effects on comprehension and word inference ability. *The Modern Language Journal*, 73(2), 135-149.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Knight, S. L., Padron, Y. N., & Waxman, H. C. (1985). The cognitive reading strategies of ESL students. *TESOL Quarterly*, 19(4), 789-792.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, 19(2), 193-220.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. New York: Cambridge University Press.

Kunnan, A. J. (1995). *Test-taker characteristics and performance: A structural equation modeling approach*. Cambridge: Cambridge University Press.

Kunnan, A. J. (1998). An introduction to structural equation modeling for language assessment research. *Language Testing*, 15(3), 295-332.

Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals*, 18(5), 405-411.

Lee, J. F., & Musumeci, D. (1988). On hierarchies of reading skills and text types. *The Modern Language Journal*, 72(2), 173-183.

Lee, J.-W., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly*, 31(4), 713-739.

Macaro, E. (2004). Fourteen features of a language learner strategy. Working Paper 4.

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *The Modern Language Journal*, 90(3), 320-337.

Macaro, E., & Erler, L. (2008). Raising the achievement of young-beginner readers of French through strategy instruction. *Applied Linguistics*, 29(1), 90-119.

Marsh, H., & Balla, J. (1994). Goodness of fit in confirmatory factor analysis: The effect of sample size and model parsimony. *Quality and Quantity: International Journal of Methodology*, 28(2), 185-217.

Marsh, H., Balla, J., & Hau, K. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides, & R. E. Schumaker (Eds.), *Advanced structural equation modeling: Issues and techniques*. Hillsdale, NJ: Erlbaum, 315-353.

Marsh, H., Balla, J., & McDonough, R. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391-410.

McDonough, S. H. (1995). *Strategy and skill in learning foreign language*. London:

Arnold.

McLeod, B., & McLaughlin, B. (1986). Restructuring or automaticity? Reading in a second language. *Language Learning*, 36(2), 109-123.

Mertens, D. M. (1998). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches*. Thousand Oaks, CA: Sage.

Ministry of Education (2006). *Senior high school curriculum outlines*. Retrieved June 28, 2006. http://www.edu.tw/EDU_WEB/EDU_MGT/HIGH-SCHOOL/EDU2890001/i1301/course/current/subject_3.htm?Open

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94(2), 249-259.

Mokhtari, K., & Reichard, C. A. (2004). Investigating the strategic reading processes of first and second language readers in two different cultural contexts. *System*, 32(3), 379-394.

Morrison, K. R. B. (1993). *Planning and accomplishing school-centered evaluation*. Norfolk: Peter Francis Publications.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus: The comprehensive modeling program for applied researchers*. Los Angeles, CA: Muthén & Muthén.

Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journal*, 89(1), 76-91.

Nassaji, H. (2002). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning*, 52(2), 439-481.

Nassaji, H. (2003a). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4), 645-670.

Nassaji, H. (2003b). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, 87(2), 261-276.

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, 6(2), 199-215.

Nieh, P.-L. (2003). *Elementary graduates' EFL proficiency in Taiwan*. Unpublished doctoral dissertation, National Kaohsiung Normal University, Taiwan.

Nikolov, M. (2006). Test-taking strategies of 12- and 13-year-old Hungarian learners of EFL: Why whales have migraines. *Language Learning*, 56(1), 1-51.

Nunnally, J. C., & Bernstein, I. J. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Pinter Publishers.

Oxford, R. (1996). Employing a questionnaire to assess the use of language learner strategies. *Applied Language Learning*, 7(1&2), 25-45.

Oxford, R., Cho, Y., Leung, S., & Kim, H. (2004). Effect of the presence and difficulty of task on strategy use: An exploratory study. *International Review of Applied Linguistics in Language Teaching*, 42(1), 1-47.

Oxford, R., & Cohen, A. D. (1992). Language learning strategies: Crucial issues of concepts and classification. *Applied Language Learning*, 3(1&2), 1-35.

Oxford, R., & Nyiko, M. (1989). Variables affecting choice of language learning strategies by university students. *The Modern Language Journal*, 73(2), 291-300.

Padron, Y. N., & Waxman, H. C. (1988). The effects of ESL students' perceptions of their cognitive strategies on reading achievement. *TESOL Quarterly*, 22(1), 146-150.

Paribakht, T. S. (2004). The role of grammar in second language lexical processing. *RELC Journal*, 35(2), 149-160.

Parry, K. (1996). Culture, literacy, and L2 reading. *TESOL Quarterly*, 30(4), 665-692.

Perkins, K., Brutten, S. R., & Pohlmann, J. T. (1989). First and second language reading comprehension. *RELC Journal*, 20(1), 1-9.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26-56.

Phakiti, A. (2008). Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests. *Language Testing*, 25(2), 237-272.

Pichette, F., Segalowitz, N., & Connors, K. (2003). Impact of maintaining L1 reading skills on L2 reading skills development in adults: Evidence from speakers of Serbo-Croatian learning French. *The Modern Language Journal*, 87(3), 391-403.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies.

Reading Research Quarterly, 15(4), 273-294.

Purpura, J. E. (1997). An analysis of the relationships between test-takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 47(2), 289-325.

Purpura, J. E. (1998a). The development and construct validation of an instrument designed to investigate selected cognitive background characteristics of test-takers. In A. J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing research colloquium, long beach*. London: Lawrence Erlbaum Associates, 111-139.

Purpura, J. E. (1998b). Investigating the effects of strategy use and second language test performance with high- and low-ability test takers: A structural equation modeling approach. *Language Testing*, 15(3), 333-379.

Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach.* Cambridge: Cambridge University Press.

Raykov, T. (2001). Approximate confidence interval for difference in fit to structural equation models. *Structural Equation Modeling*, 8(3), 458-469.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Ridgway, T. (1997). Thresholds of the background knowledge effect in foreign language reading. *Reading in a Foreign Language*, 11(1), 151-168.

Ritter, S. A., & Idol-Maestas, L. (1986). Teaching middle school students to use a test-taking strategy. *Journal of Educational Research*, 79(6), 350-357.

Robson, C. (1993). *Real world research: A resource for social scientists and practitioner-researchers*. Oxford: Blackwell.

Rogers, W. T., & Bateson, D. J. (1991). Verification of a model of test-taking behavior of high school seniors. *Journal of Experimental Education*, 59(3), 331-349.

Rogers, W. T., & Bateson, D. J. (1994). Verification of a model of test-taking behavior of high school seniors. *Alberta Journal of Educational Research*, 15(2), 195-211.

Rosenblatt, L. M. (1978). *The reader: The text: The poem*. Carbondale, IL: Southern Illinois University Press.

Rubin, J. (1981). The study of cognitive processes in second language learning. *Applied Linguistics*, 2(2), 117-133.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance* VI. Hillsdale, NJ: Erlbaum, 573-603.

Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441-474.

Samuels, S. J., & Kamil, M. L. (1988). Models of the reading process. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading*. Cambridge: Cambridge University Press, 22-36.

Sarig, G. (1987). High level reading in the first and the foreign language: Some comparative process data. In P. L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Research in reading English as a second language*. Washington, DC: TESOL, 107-120.

Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning*, 43(3), 313-344.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press, 199-277.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning*, 48(1), 71-106.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351-371.

Schuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research*, 56(4), 411-436.

Schulz, R. A. (1983). From word to meaning: Foreign language reading instruction after the elementary course. *The Modern Language Journal*, 67(2), 127-134.

Schumacker, R. E., & Lomax, R. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Seliger, W. H., & Shohamy, E. (2000). *Second language research methods*. Oxford: Oxford University Press.

Shen, M.-Y. (2004). *Implementing context-based instruction for word inferencing in EFL reading: The effects on technical college learners*. Unpublished doctoral dissertation, Tamkang University, Taiwan.

Sheorey, R. (1999). An examination of language learning strategy use in the setting of an indigenized variety of English. *System*, 27(2), 173-190.

Sheorey, R., & Edit, S. (2004). Perceptions of reading strategy use among Hungarian EFL students. *Research News: The Newsletter of the IATEFL Research SIG*, 14, 19–23.

Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System*, 29(4), 431-449.

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, 24(1), 99-128.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Smith, F. (1982). *Understanding reading* (3rd ed.). New York: Holt, Rinehart & Winston.

Stanvich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16(1), 32-71.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stevenson, M., Schoonen, R., & de Glopper, K. (2003). Inhibition or compensation? A multidimensional comparison of reading processes in Dutch and English. *Language Learning*, 53(4), 765-815.

Swaffar, J. K. (1988). Readers, texts, and second languages: The interactive processes. *The Modern Language Journal*, 72(2), 123-149.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins Publishers.

Taillefer, G. F. (1996). L2 reading ability: Further insight into the short-circuit hypothesis. *The Modern Language Journal*, 80(4), 461-477.

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78-102.

Upton, T. A., & Lee-Thompson, L. (2001). The role of the first language in second language reading. *Studies in Second Language Acquisition*, 23(4), 469-495.

Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice*. London: Longman.

Usó-Juan, E (2006). The compensatory nature of discipline-related knowledge and

English-language proficiency in reading English for Academic purposes. *The Modern Language Journal*, 90(2), 210-227.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Steveson, M. (2004). Linguistic knowledge, processing speed and metacognitive knowledge in first and second language reading comprehension: A componential analysis. *Journal of Educational Psychology*, 96(1), 19-30.

Vandergrift, L., Goh, C. C. M., Marechal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning*, 56(3), 431-462.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, 25(3), 315-339.

White, C., Schramm, K., & Chamot, A. U. (2007). Research methods in strategy research: Re-examining the toolbox. In A. D. Cohen, & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice*. Oxford: Oxford University Press, 93-116.

Wiersma, W., & Jurs, G. (2005). *Research methods in education: An introduction* (8th ed.). Boston, MA: Allyn & Bacon.

Williams, E., & Moran, C. (1989). Reading in a foreign language at intermediate and advanced levels with particular reference to English. *Language Teaching*, 22(4), 217-228.

Winograd, P., & Hare, V. C. (1988). Direct instruction of reading comprehension strategies: The nature of teacher explanation. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction and evaluation*. San Diego, CA: Academic Press, 121-139.

Wolf, D. F. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *The Modern Language Journal*, 77(4), 473-489.

Wolff, D. (1987). Some assumptions about second language text comprehension. *Studies in Second language Acquisition*, 9(3), 307-326.

Wu, M.-L., & Tu, G.-T. (2005). *SPSS & the application and analysis of statistics* (2nd ed.). Taipei: Wu-Nan Press.

吳明隆&涂金堂 (民 94)。SPSS 與統計應用分析（第二版）。台北: 五南圖書。

Yamashita, J. (2002). Reading strategies in L1 and L2: Comparison of four groups of readers with different reading ability in L1 and L2. *ITL: Review of Applied Linguistics*, 135-136, 1-35.

Yang, Y.-F. (2002). Reassessing readers' comprehension monitoring. *Reading in a Foreign Language*, 14(1), 18-42.

Yang, Y.-F. (2006). Reading strategies or comprehension monitoring strategies? *Reading Psychology*, 27(4), 313-343.

Yang, X., & Zhang, W. (2002). Yuanrenzhi yu Zhongguo daxuesheng Yingyu yuedu lijie xiangguan yanjiu [The correlation between metacognition and EFL reading comprehension of Chinese college students]. *Foreign Language Teaching and Research*, 34(3), 213-218.

Yiğter, K., Sariçoban, A., & Gürses, T. (2005). Reading strategies employed by ELT learners at the advanced level. *The Reading Matrix*, 5(1), 124-139.

Yorio, C. (1971). Some sources of reading problems for foreign language learners. *Language Learning*, 21(1), 107-115.

Appendix 1

An English Language Knowledge Test used in the Current Study

An English Language Knowledge Test

Purpose: This English language knowledge test is to understand your English language knowledge. You are expected to take the test as carefully as you do in normal test situations. Your contribution will be appreciated. The test result will be treated as strictly confidential. Contact me if you want to know the result of the test. E-mail: www.weitsung@soton.ac.uk 目的：此英語語文測驗主要在了解你的英語知識範圍。希望你能把它視為正式的考試，仔細作答。測驗的結果將會嚴格保密，請你放心作答。如果你想知道測驗的結果，請 E-mail 給本人。
Instruction: This English language knowledge test consists of a grammar subtest and a vocabulary subtest. A grammar subtest contains 29 test items, while a vocabulary subtest contains 27 test items. Except the first section of the vocabulary subtest, each test item constitutes 4 options. Read the question and choose an appropriate answer. You have 20 minutes for the grammar subtest and 25 minutes for the vocabulary subtest. Please write your answer on the answer sheet. 測驗說明：此英語語文測驗包含文法測驗及單字測驗。文法測驗有二十九道題目，單字測驗有二十七道題目。除了單字測驗的第一大題之外，每個題目都有四個選項。看完題目後，請選出一個適當的答案。文法測驗作答時間為二十分鐘；單字測驗作答時間為二十五分鐘。請把答案寫在答案紙上。

I. A Grammatical Subtest (文法測驗) (choose an appropriate answer) (請選出一個適當的答案)

1. _____ children like to eat chocolate.
   (A) Most (B) Most of (C) The most (D) For most
2. John is _____ tennis player I have ever seen.
   (A) a good (B) a better (C) the better (D) the best
3. I will never forget _____ the A-li Mountain for the first time.
   (A) see (B) seeing (C) for seeing (D) to have seen
4. _____ is no knowing what will happen tomorrow.
   (A) It (B) One (C) That (D) There
5. _____ her work, Susan took a rest under a tree and felt happy.
   (A) Finished (B) Had finished (C) Having finished (D) She finished
6. Under the table in this room _____ several books.
   (A) is (B) are (C) has (D) have
7. A: When are you moving into a new house?
   B: Maybe _____ the fifth of June.
   (A) on (B) in (C) at (D) for
8. A: John, you did not tell us to have a test today?
   B: Yes, I did. I remember _____ you last Monday.
   (A) tell (B) told (C) telling (D) to have told
9. A: Do you still study Japanese?
   B: Oh, yes. I _____ it since I graduated from high school.
   (A) study (B) am studying (C) have been studied (D) have been studying
10. A: Did you watch any of this famous actor's films?
    B: No, I don't like _____ he plays.
    (A) them (B) such (C) which (D) what
11. A: Did you hear that Amy did poorly in the math exam?
    B: Yes. If she _____ harder, she wouldn't have.
    (A) studied (B) would study (C) had studied (D) had been studied
12. A: Did Kevin go to the movie last night?
    B: Yes, but he _____ home to take care of his little brother.
    (A) stays (B) has stayed (C) should stay (D) should have stayed

13. A: This math question is really hard!

B: Yeah, no one in class knows _____ to answer it.

(A) which (B) such (C) that (D) how

14. A: Has John seen the doctor yet?

B: Yes, and she suggested that he _____ smoking.

(A) stop (B) stops (C) stopped (D) had stopped

15. A: You look tired. Why?

B: My mom had me _____ my room and the living room.

(A) to clean (B) cleaning (C) cleaned (D) clean

16. A: What are we going to talk about at today's meeting?

B: Maybe we are going on to talk about the problem _____ at the last meeting.

(A) discussing (B) discussed (C) was discussed (D) had been discussed

17. A: Do you know Jane needs money badly?

B: Yes. I wish I _____ her some.

(A) lend (B) will lend (C) could lend (D) had lent

18. A: What are you going to do this weekend?

B: If it _____ , I'll go shopping with my friend.

(A) doesn't rain (B) won't rain (C) isn't raining (D) won't be raining

19. A: Why is Sophie so angry?

B: Because her husband kept her ______ at the restaurant for over two hours.

(A) wait (B) waiting (C) waited (D) to wait

20. A: I was told that you are going to marry a rich man next week.

B: ______ told you, that is not true.

(A) Who (B) Someone (C) What (D) Whoever

21. A: Oh! It is 7:50. You may miss the train.

B: Yes, I will miss the train _____ I hurry.

(A) then (B) hence (C) besides (D) unless

22. A: What's wrong with Jack?

B: He went away sadly without a word _____ .

(A) spoken (B) speaking (C) to speak (D) to be spoken

23. A: May I invite Tim and Grace to my birthday party?

B: Of course, you can invite _____ you like.

(A) that (B) why (C) however (D) whomever

24. A: How long have you lived in Taiwan?

B: By next July, I _____ in Taiwan for five years.

(A) would have lived (B) will have lived (C) will be living (D) will live

25. A: I called you up about 8:00 last night, but no one answered the phone.

B: I _____ a shower when you called.

(A) take (B) was taking (C) will take (D) had taken

26. A: Why are you so sad?

B: My girlfriend told me that she ______ to France next week.

(A) will go (B) will have gone (C) would go (D) would have gone

27. A: Mary isn't rich, right?

B: No. However, she always lives _____ she were a rich woman.

(A) therefore (B) as if (C) by means of (D) in order to

28. A: Rose quit her part-time job.

B: Did she say _____ ?

(A) why she quit her job (B) why did she quit her job

(C) why her job she quit (D) why did her job she quit

29. A: Do you know anything about our new math teacher?

B: She is a PhD student ______ near our school.

(A) whom she lives with (B) which lives (C) she lives (D) that lives

PLEASE STOP AND WAIT FOR FURTHER INSTRUCTIONS

II. A Vocabulary Subtest (單字測驗)

Part I. Matching Items (next to each word, write the number of its meaning)
(配合題，填寫代號即可)

Section A.

1. taste	(A) to do things in a particular way
2. spread	(B) to make it easier for an activity to happen
3. behave	(C) to eat or drink something to see what it is like
	(D) to become very successful or very strong and healthy
	(E) to become known about or used by more and more people

Section B

4. playground	(A) a feeling of liking or love and caring
5. treatment	(B) an action that breaks a law or an agreement
6. exhibition	(C) something that is done to cure someone who is ill
7. intelligence	(D) the ability to learn, understand, and think about things
8. hazard	(E) an area for children to play, especially at a school or in a park
	(F) something that may be dangerous, or cause accidents or problems
	(G) a show of painting, photos or other interesting things that people can go to see

Section C.

9. alone	(A) very big, impressive or beautiful
10. sensitive	(B) without any friends or people you know
11. voluntary	(C) done willingly and without being forced
	(D) very weak, especially because you are old or ill
	(E) able to understand other people’s feelings and problems

Part II. Sentence Completion (choose an appropriate answer) (請選出一個適當的答案)

12. The boy was hurt badly in the car accident and died ______ afterwards.

(A) shortly (B) mostly (C) easily (D) hardly

13. As computers are getting less expensive, they are ______ used in schools and offices today.

(A) totally (B) chiefly (C) rarely (D) widely

14. I called the airline to _____ my flight a week before I left England.

(A) explain (B) confirm (C) attack (D) strike

15. Miss Chang always tries to answer all questions from her students. She will not ______ any of them even if they may sound stupid.

(A) ignore (B) reduce (C) arrest (D) locate

16. All the train service to and from Taipei were _____ because of the heavy thunderstorm.

(A) benefited (B) cancelled (C) debated (D) advised

17. The woman told the truth to her lawyer without ______ because he was the only person she could depend on.

(A) foundation (B) occupation (C) reservation (D) combination

18. If you want to borrow magazines, tapes, or CDs, you can visit the library. They are all _____ there.

(A) marvelous (B) available (C) sufficient (D) impressive

19. To avoid being misled by news reports, we should learn to _____ between facts and opinions.

(A) suppose (B) distinguish (C) negotiate (D) complicate

20. If you exercise regularly, your blood _____ will be improved and you will feel more healthy.

(A) circulation (B) landscape (C) harmony (D) assistance

21. In order to write a report on stars, we decided to _____ the stairs in the sky every night.

(A) define (B) sprinkle (C) observe (D) frustrate

22. Irene does not throw away used envelopes. She _____ them by using them for taking telephone messages.

(A) isolates (B) disguises (C) recycles (D) manufactures

23. Our team will certainly win this baseball game, because all the players are highly _____ .

(A) motivated (B) dominated (C) estimated (D) illustrated

24. Your desk is crowded with many unnecessary things. You have to ______ some of them.

(A) remove (B) renew (C) resist (D) remain

25. Helen _____ with anger when she saw her boyfriend kissing another girl.

(A) relaxed (B) collided (C) defeated (D) exploded

26. Jack doesn't look _____ , but he is, in fact, good at sports, especially baseball.

(A) graceful (B) athletic (C) unique (D) conservative

27. Anne feared giving a speech before three hundred people; even thinking about it made her ______ .

(A) anxious (B) passionate (C) optimistic (D) sorrowful

The End

Appendix 2

A Reading and Test-taking Strategy Use Questionnaire used in the Current Study

A Strategy Use Questionnaire

Purpose: This questionnaire aims to understand what you do and how you tackle a multiple-choice EFL reading comprehension test. This is not a test; therefore, there are no "right" or "wrong" answers. The result of the questionnaire is irrelevant to your academic records. However, you are still expected to fill in the questionnaire carefully as well as honestly, and your contribution will be appreciated. The information you provide will be treated as strictly confidential. Contact me if you have any questions. E-mail：www.weitsung@soton.ac.uk. Thank you very much for your help.
Direction: Recall what you did and how you did it as you were taking a multiple-choice EFL reading comprehension test. Read the following statement and see how far these statements match your strategy use. Circle 5 (strongly agree), 4 (agree), 3 (partly agree), 2 (partly disagree), 1 (disagree) and 0 (strongly disagree). After you have finished this questionnaire, please check it again to make sure that you respond to each statement. You have twenty-five minutes to respond to this questionnaire.

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
1. When I got the test, I knew what I was going to do first.	5	4	3	2	1	0
2. When I took the test, I tried to read the passage roughly for a general understanding.	5	4	3	2	1	0
3. When I took the test, I tried to use clues from test questions to decide whether to read a particular part of the passage.	5	4	3	2	1	0
4. When I took the test, I tried to read the passage quickly for particular information.	5	4	3	2	1	0
5. When I read the passage, I tried to translate a word into Chinese.	5	4	3	2	1	0
6. During the reading process, I was aware that I did not understand the meaning of a word.	5	4	3	2	1	0
7. When I encountered an unknown word, I tried to mark it.	5	4	3	2	1	0
8. When I encountered an unknown word, I tried to guess its meaning by breaking it into parts.	5	4	3	2	1	0
9. When I encountered an unknown word, I tried to guess its meaning by using context clues.	5	4	3	2	1	0
10. When I encountered an unknown word, I tried to infer its meaning by using the clues from test questions.	5	4	3	2	1	0
11. During the reading process, I tried to infer the meaning of an unknown word from the immediate sentence.	5	4	3	2	1	0

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
12. During the reading process, I tried to translate the whole sentence into Chinese.	5	4	3	2	1	0
13. During the reading process, I tried to identify key words in the sentence.	5	4	3	2	1	0
14. During the reading process, I tried to substitute a word in the sentence to help me understand the meaning of the sentence.	5	4	3	2	1	0
15. During the reading process, I tried to use my words to interpret the meaning of the sentence.	5	4	3	2	1	0
16. During the reading process, I tried to make an inference about the sentence I read.	5	4	3	2	1	0
17. During the reading process, I tried to question myself whether I understood the meaning of the sentence I read.	5	4	3	2	1	0
18. During the reading process, I tried to use grammar rules to understand the meaning of the sentence I read.	5	4	3	2	1	0
19. During the reading process, I tried to identify the importance of the sentence I read.	5	4	3	2	1	0
20. When I did not understand the meaning of a sentence, I tried to reread it.	5	4	3	2	1	0
21. When I read a sentence, I thought whether it was related to test questions.	5	4	3	2	1	0
22. When I read a sentence, I noticed it was related to test questions.	5	4	3	2	1	0

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
23. During the reading process, I tried to associate something else with the sentence I read.	5	4	3	2	1	0
24. When I read the passage, I tried to marked the sentence that I did not understand.	5	4	3	2	1	0
25. During the reading process, I was aware that I roughly understood the meaning of the sentence although there was a word I did not understand.	5	4	3	2	1	0
26. During the test-taking process, I read the relevant information about a test question and immediately answered it.	5	4	3	2	1	0
27. When I read the passage, I tried to predict what I was going to read.	5	4	3	2	1	0
28. When I read the passage, I tried to check if my inference was correct.	5	4	3	2	1	0
29. When I read the passage, I tried to summarize what I read.	5	4	3	2	1	0
30. When I read a paragraph, I tried to refer to the previous paragraph to better understand what I read.	5	4	3	2	1	0
31. When I read the passage, I tried to integrate the information from different parts of the passage.	5	4	3	2	1	0
32. When I read the passage, I tried to use what I already knew to help me understand the passage.	5	4	3	2	1	0
33. When I did not understand a part of the passage, I tried to get clues from test questions to help me understand it.	5	4	3	2	1	0

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
34. When I read the passage, I tried to identify the important and the less important parts of the passage.	5	4	3	2	1	0
35. When I read the passage, I tried to mark key points in the passage.	5	4	3	2	1	0
36. When I read the passage, I tried to remember where key points were in the passage.	5	4	3	2	1	0
37. When I read the passage, I tried to skip confusing parts of the passage, e.g., time or people's names.	5	4	3	2	1	0
38. When I read the passage, I tried to ask myself questions about what I read.	5	4	3	2	1	0
39. When I read the passage, I tried to relate it to my personal experiences.	5	4	3	2	1	0
40. When I read the passage, I tried to respond to the content of the passage with my personal opinions.	5	4	3	2	1	0
41. When I read the passage, I tried to respond to the content of the passage with my personal feelings.	5	4	3	2	1	0
42. When I read the passage, I tried to have a picture in mind about what I read.	5	4	3	2	1	0
43. When I read the passage, I had test questions in mind.	5	4	3	2	1	0
44. When I read the passage, I tried to predict that some key points would become test questions.	5	4	3	2	1	0
45. During the reading process, I was aware that I understood a part of the passage.	5	4	3	2	1	0

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
46. During the reading process, I was aware that I did not understand a part of the passage.	5	4	3	2	1	0
47. During the reading process, I knew that I didn't concentrate.	5	4	3	2	1	0
48. When I read the passage, I was aware of the difficulty of the passage.	5	4	3	2	1	0
49. When I read the passage, I knew my weaknesses in reading.	5	4	3	2	1	0
50. When I did not understand what I read, I tried to read it slowly.	5	4	3	2	1	0
51. When I did not understand the paragraph, I tried to reread it.	5	4	3	2	1	0
52. During the test-taking process, I was aware of what I did.	5	4	3	2	1	0
53. During the test-taking process, I was aware of how was done.	5	4	3	2	1	0
54. During the test-taking process, I was aware of which strategy was used in answering different types of test questions.	5	4	3	2	1	0
55. When I answered test questions, I tried to recall a part of the passage.	5	4	3	2	1	0
56. During the question-answering process, I tried to understand the meanings of test questions appropriately.	5	4	3	2	1	0
57. When I answered test questions, I tried to answer them in different orders based on their difficulty.	5	4	3	2	1	0
58. When I did not get an answer to a test question, I tried to skip it and return to it later.	5	4	3	2	1	0

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
59. When I answered test questions, I tried to find a related paragraph by using clues from test questions.	5	4	3	2	1	0
60. When I answered test questions, I tried to get my answers based on my understanding of the passage.	5	4	3	2	1	0
61. During the test-taking process, I got my answers even though I roughly understood the passage.	5	4	3	2	1	0
62. When I answered test questions, I selected an option through reasoning.	5	4	3	2	1	0
63. When I answered test questions, I tried to match options with a part of the passage.	5	4	3	2	1	0
64. When I answered test questions, I tried to mark the differences among options.	5	4	3	2	1	0
65. When I answered test questions, I selected an option because the others seemed unreasonable.	5	4	3	2	1	0
66. During the test-taking process, I was aware that I did not understand options.	5	4	3	2	1	0
67. When I answered test questions, I had confidence in the answer I chose.	5	4	3	2	1	0
68. When I answered test questions, I tried to spend more time on difficult test questions.	5	4	3	2	1	0
69. When I answered test questions, I was ready to change an answer if necessary.	5	4	3	2	1	0
70. I noticed how much time I still had when I took the test.	5	4	3	2	1	0
71. I tried to finish the test as soon as possible during the test-taking process.	5	4	3	2	1	0

72. During the test-taking process, I tried to double-checked the answers.	5	4	3	2	1	0

III. Please write down the basic information about yourself.

1. School:
2. Class:
3. Number:
4. Gender: **A. Male** **B. Female**
5. How many years have you learned English?

   A. Under five years. **B. Five to ten years.** **C. Above ten years.**

6. Have you ever read the passages in the test before?

   Yes. **No.**

   6.1 If "**yes**", which reading passage have you read before?

   A B C D E F

7. Were you familiar with the topic of the reading passages?

   Yes. **No.**

   7.1 If "**yes**", which reading passage were you familiar with?

   A B C D E F

8. Have you ever learned English in cram schools?

   Yes. **No.**

9. In terms of enhancing your English reading comprehension test performance, please rate the importance of the following items.

		Very important	Important	Partly important	Less important	Least important
9.1	Having enough vocabulary knowledge	**5**	**4**	**3**	**2**	**1**
9.2	Having enough grammatical knowledge	**5**	**4**	**3**	**2**	**1**
9.3	Being aware of some reading strategies	**5**	**4**	**3**	**2**	**1**
9.4	Being aware of how to use reading strategies appropriately	**5**	**4**	**3**	**2**	**1**
9.5	Being aware of some test-taking strategies	**5**	**4**	**3**	**2**	**1**
9.6	Being aware of how to use test-taking strategies appropriately	**5**	**4**	**3**	**2**	**1**

10. Please self-rate your English ability in terms of four language skills from 1 to 5 (1=poor, 5=excellent).

10.1	Listening	**5**	**4**	**3**	**2**	**1**
10.2	Speaking	**5**	**4**	**3**	**2**	**1**
10.3	Reading	**5**	**4**	**3**	**2**	**1**
10.4	Writing	**5**	**4**	**3**	**2**	**1**

Thanks for your cooperation!

Appendix 3

A Reading Comprehension Test used in the Current Study

A Reading Comprehension Test

Purpose: This reading test aims to understand your English reading ability. You are expected to take the test as carefully as you do in normal test situations. Your contribution will be appreciated. The test result will be treated as strictly confidential. Contact me if you want to know the result of the test. E-mail: www.weitsung@soton.ac.uk 目的：此閱讀測驗在了解你的英文閱讀能力。希望你能把它視為正式的考試，仔細地作答。測驗的結果將會嚴格保密，請你放心作答。如果你想知道測驗的結果，請 **E-mail** 給本人。E-mail：www.weitsung@soton.ac.uk。
Instruction: In this test, there are six reading passages and seventeen questions. Each reading passage is followed by two to five questions. After you read the passage, please answer its following questions. You have forty-five minutes to complete this test. Please write your answer on the answer sheet. 測驗說明：此閱讀測驗包含六篇文章和十七題問題。每篇文章有二至五題問題。看完文章後，請回答其下列問題。作答時間為四十五分鐘。請把答案寫在答案紙上。

Reading Passage A

I usually go to work by subway, and I get to work by 8:00 A.M. Before I start my job, I put on my uniform and look at myself in the mirror to make sure that I look neat. At 8:30 in the morning, I go on duty. I usually eat lunch from twelve to one and generally take a fifteen-minute break in the morning and in the afternoon. At the 4:30 in the afternoon, I go off duty.

I enjoy my job very much. I meet all kinds of people and talk to everyone. Many people ask me questions, and I give them the necessary information. I try to be very helpful. I always call out floors very clearly, and I am constantly on the move. Most men take off their hats in my car, and sometimes I have to tell passengers to put out their cigarettes. Some people smile at me, but others just ignore me. In fact, my life can be described as consisting of a series of “ups” and “downs.”

1. The passage is written mainly to describe ______ .

   (A) what kinds of people the author works with
   (B) what a typical day is like for the author
   (C) what “life” means to the author
   (D) what the author looks like

2. The expression “constantly on the move” in the passage refers to the fact that _____ .

   (A) the author frequently helps passengers move their baggage
   (B) the author meets all kinds of people and talks to everyone
   (C) the author seldom stays in one place for a long while
   (D) the author always calls out floors very clearly

Reading Passage B

Sometimes the real world can be a confusing place. It is not always fair or kind. And in the real world there are not always happy endings. That is why, every once in a while, we like to escape into the world of fantasy – a place where things always go our way and there is always a happy ending.

We want to believe in fantastic creatures in imaginary lands. We want to believe in magic powers, good friends, and the power of good to overcome evil. We all fantasize about being able to fly and lift buildings off the ground. And how good a magic sword would feel in our hand as we go off to kill a dragon or win the hand of a beautiful princess.

The amazing adventures of Superman, Peter Pan, and Harry Potter have charmed many people, children and adults alike. The main reason is that these stories offer us chances to get away from this real, frustrating world and allow us to find some magical solutions to our problems. For example, Superman always arrives in the nick of time to prevent a disaster from happening, Peter Pan can fly at will to tease the bad guy Captain Hook, and Harry Potter has his magic power to take revenge on his uncle, aunt and cousin, who always ill-treat him.

3. This article about fantasy literature is intended to _____ .

(A) explain why people like to read it
(B) laugh at those people reading it
(C) criticize its unrealistic concepts
(D) teach people to avoid disasters

4. People enter the world of fantasy for the following reasons **EXCEPT** that ______ .

(A) we can always have our wishes fulfilled
(B) the real world is often disappointing
(C) we can find happy endings there
(D) the world of fantasy frightens us

Reading Passage C

A sense of humor is just one of the many things shared by Alfred and Anthony Melillo, 64-year-old twin brothers from East Haven who made history in February 2002. On Christmas Eve, 1992, Anthony had a heart transplant from a 21-year-old donor. Two days before Valentine's Day in 2002, Alfred received a 19-year-old heart, marking the first time on record that twin adults each received heart transplants.

"**I'm 15 minutes older than him, but now I'm younger because of my heart** and I'm not going to respect him," Alfred said with a grin, pointing to his brother while talking to a roomful of reporters, who laughed frequently at their jokes.

While the twins knew that genetics might have played a role in their condition, they recognized that their eating habits might have also contributed to their heart problems. "We'd put half a pound of butter on a steak. I overdid it on all the food that tasted good, so I guess I deserved what I got for not dieting properly."

The discussion moved to Anthony's recovery. In the five years since his heart transplant, he had been on an exercise program where he regularly rode a bicycle for five miles, swam each day, and walked a couple of miles. He was still on medication, but not nearly as much as Alfred, who was just in the early stage of his recovery.

"Right now I feel pretty young and I'm doing very well," Anthony said, "I feel like a new person." Alfred said his goal, of course, was to feel even better than his brother. But, he added, "I love my brother very much. We're very close and I'm sure we'll do just fine."

5. What did Alfred and Anthony think caused their heart problems?

(A) Diet.
(B) Exercise.
(C) Surgery.
(D) Medicine.

6. Why did Alfred say, "**I'm 15 minutes older than him, but now I'm younger because of my heart**"?

(A) His heart transplant surgery was more successful than Anthony's.
(B) His recovery from the heart surgery was faster than Anthony's.
(C) His exercise program was better than Anthony's.
(D) His new heart was younger than Anthony's.

Reading Passage D

On December 26, 2003, the worst earthquake in more than a decade devastated Bam, a historic city in Iran. At least 25,000 people died in the quake – nearly a third of the city's population. And thousands more were left homeless, hungry, and grieving.

Bam was a city of mud-brick houses, old monuments and an ancient castle. But nearly everything crumbled in the disaster. One reason the earthquake caused such damage was that Bam's buildings were made mostly from baked mud. These buildings collapsed in heaps of dust and sand.

Bam was best known for its 2,000-year-old castle built out of mud, straw, and the trunks of palm trees. The castle was so big that it was once the city of Bam itself. Public dwellings lined its ground level; a marketplace and two mosques also fit comfortably inside.

Bam once blossomed as a trading post on the Silk Road. In the 16^{th} and 17^{th} centuries, treasures from the Far East were carried along the road into the capital cities of Europe. Fifty years ago, teams of architects began restoring the historic treasures of the city. Even since, thousands of visitors have come to admire them.

In the face of this tragedy, food and other supplies from around the world landed in the provincial capital of Kerman on Sunday. With such support, spiritual leader Ayatollah Ali Khamenei vowed, "We will rebuild Bam stronger than before."

7. What was Bam most famous for?

(A) An old mud and straw castle.
(B) Treasures from the Far East.
(C) Frequent earthquakes.
(D) Beautiful palm trees.

8. The use of baked mud for buildings explains ______ .

(A) why the earthquake caused such damage
(B) why Bam developed into a trading post
(C) why Bam collected so many treasures
(D) why the earthquake struck Bam

9. Which of the following is TRUE about the earthquake in Bam?

(A) The city of Bam would be deserted after the earthquake.
(B) The 2003 earthquake was the first one in its history.
(C) Not many countries sent food and supplies to Bam.
(D) About 50,000 people survived the earthquake.

Reading Passage E

Today's teen consumer market is the most profitable it has ever been. Even though 65% of teens claim that they rely on themselves for their fashion ideas, it is estimated that less than 20% of the teen population is innovative enough to drive fashion trends, according to a recent study by a marketing firm. Marketers recognize this fact and often use elements of youth culture to promote their products. Perhaps one of the best examples is their use of hip-hop culture. It is reported that hip-hop fashion alone generates $750 million to $1 billion annually. Sales of rap music and videos each exceed that amount.

Rap's rise and sustained global popularity is a good illustration of how influential youth culture is on youth attitudes and behavior. Remember when Madonna hit the charts with her bra in full view while singing about "virginity"? Soon after that, adolescent girls around the world began wearing their underwear outside their clothes.

Fashion designer Tommy Hilfiger was fully aware of the power of youth culture. He marketed his brand by giving clothes to famous MTV stars and featuring teen stars in his print ads. Picking up on teens' interest in computer games, Hilfiger sponsored a Nintendo competition and installed Nintendo terminals in his stores. The payoff? Teens rated Hilfiger jeans as their number one brand in a survey in 2000.

10. What is the best title for this passage?

(A) The Power of Youth Culture
(B) The Importance of Marketing
(C) The Success of Tommy Hilfiger
(D) The Popularity of Hip-hop Fashion

11. How much money do sales of rap music and videos together make each year?

(A) Between $750 million and $1 billion.
(B) Between $500 million and $750 million.
(C) More than $1 billion.
(D) Less than $500 Million.

12. According to the passage, which of the following statements is true?

(A) Marketers recognize youth culture as a part of hip-hop culture.
(B) Madonna led the fashion of wearing underwear outside clothes.
(C) Many teenagers make a lot of profits in the fashion market today.
(D) The purchasing power of teenagers has been decreasing over the years.

Reading Passage F

Joy Hirsch, a neuroscientist in New York, has recently found evidence that children and adults don't use the same parts of the brain when learning a second language. He used an instrument called an MRI (magnetic resonance imaging) to study the brains of two groups of bilingual people. One group consisted of those who had learned a second language as children. The other consisted of people who learned their second language later in life. People from both groups were placed inside the MRI scanner. This allowed Hirsch to see which parts of the brain were getting more blood and were more active. He asked people from both groups to think about what they had done the day before, first in one language and then the other. They couldn't speak out loud, because any movement would disrupt the scanning.

Hirsch looked specifically at two language centers in the brain – Broca's area, believed to control speech production, and Wernicke's area, thought to process meaning. He found that both groups of people used the same part of Wernicke's area no matter what language they were speaking. But how they used Broca's area was different.

People who learned a second language as children used the same region in Broca's area for both languages. People who learned a second language later in life used a special part of Broca's area for their second language – near the one activated for their native tongue.

How does Hirsch explain this difference? He believes that when language is first being programmed in young children, their brains may mix all languages into the same area. But once that programming is complete, a different part of the brain must take over a new language. Another possibility is simply that we may acquire languages differently as children than we do as adults. Hirsch thinks that mothers teach a baby to speak by using different methods such as touch, sound, and sight. And that's very different from sitting in a high school class.

13. The purpose of this passage is to ______ .

(A) describe the best ways to acquire languages at different ages
(B) describe research into the brains of bilingual people
(C) explain how to be a better second language learner
(D) explain how people become bilingual

14. In the study, the subjects were placed inside the MRI scanner to ______ .

(A) describe the best areas of the brains for learning second languages
(B) observe the movements of the brains when they spoke out loud
(C) describe the functions of the areas of the brains when they slept
(D) observe the activities of the brains when they used languages

15. The language center in the brain that is believed to control speech production is called _____ .

(A) Broca's area (B) Wernicke's area (C) native tongue (D) MRI

16. According to the passage, which of the following is TRUE for bilingual people?

(A) Those who spoke different languages always used different parts of Wernicke's area.
(B) Those who spoke the same language never used Broca's area and Wernicke's area.
(C) Those who spoke different languages always used the same part of Broca's area.
(D) Those who spoke different languages used the same part of Wernicke's area.

17. According to the passage, we can infer that ______ .

(A) unlike children, the methods that adults use to learn a second language are not different from those they use to learn their mother tongue
(B) there is a difference in the programming of a first language between children and adults
(C) Wernicke's area in our brain operates when we try to understand what other people say
(D) during the MRI scanning process, the subjects' movement contributes to the accuracy of the result

THE END

Appendix 4

Descriptive Analysis, Exploratory Factor Analysis and Confirmatory Factor Analysis

I. Introduction

This appendix primarily addresses descriptive analysis and a series of exploratory factor analyses (EFAs) with the adoption of principal components analysis[1] run on the English language knowledge test, the reading and test-taking strategy use questionnaire, as well as the multiple-choice reading comprehension test. The three measuring instruments administered in the present study were ones verified by item analyses performed in the pilot study. The purpose of descriptive analysis is to provide the information on the distributions of individual items in each measure. With regard to EFAs, the aim is to validate the three types of measures based on 834 third-graders from six senior high schools in Taiwan with a view to uncovering what components underlie these measures. These analyses also proffer a basis to construct measurement models for formulating the relationship amongst Taiwanese senior high school students' English language knowledge, strategy use, and their reading test performance. Although EFAs had been performed on these measures in the pilot study, I was interested in whether similar results would be generated from the analyses grounded on different participants. In addition, with the use of structural equation modeling (SEM) procedures, I conducted confirmatory factor analysis (CFA) on the reading and test-taking strategy use questionnaire to provide more empirical evidence for the results produced from EFAs. In the following sections, I will first focus on descriptive analysis and exploratory factor analysis for the English language knowledge test and then shift to those for the reading and test-taking strategy use questionnaire, as well as for the multiple-choice reading comprehension test.

[1] As Tabachnick and Fidell (1996) suggest, if an empirical summary of the data set is a goal, principal components analysis is an appropriate choice. Stevens (1996) also mentions a number of advantages for principal components analysis – "it is psychometrically sound, simpler mathematically and it avoids some of the potential problems with 'factor indeterminancy' associated with [other] factor analysis" (Stevens, 1996, p. 363). In addition, principal components analysis was used in Purpura's study (1998a), related to the present study, to analyze his cognitive and metacognitive strategy use questionnaires. Based on what has been stated, in the current study, I adopted principal components analysis to summarize my data set.

II. Distributions and reliabilities of the English language knowledge test

To begin with, I analyzed the item-level data stemming from the English language knowledge test, grounded on 834 third-graders from six senior high schools. The following table presents the descriptive statistics for each test item in the English language knowledge test and for the overall test.

Table 1 Distributions for test items in the English language knowledge test and for the entire test and the reliability estimate for the entire test

Variable	Mean	Std Dev	Skewness	Kurtosis	Variable	Mean	Std Dev	Skewness	Kurtosis
GQ 1	.565	.496	-.262	-1.936	***LQ 1***	.958	.201	***-4.577***	***18.993***
GQ 2	.916	.277	***-3.006***	***7.055***	LQ 2	.620	.486	-.495	-1.759
GQ 3	.728	.445	-1.026	-.951	LQ 3	.633	.482	-.553	-1.698
GQ 4	.456	.498	.178	-1.973	***LQ 4***	.958	.201	***-4.577***	***18.993***
GQ 5	.697	.460	-.857	-1.269	LQ 5	.465	.499	.140	-1.985
GQ 6	.579	.494	-.321	-1.901	LQ 6	.591	.492	-.371	-1.867
GQ 7	.687	.464	-.808	-1.350	LQ 7	.776	.417	-1.325	-.245
GQ 8	.721	.449	-.958	1.032	LQ 8	.495	.500	.019	-2.005
GQ 9	.597	.491	-.397	-1.847	***LQ 9***	.954	.209	***-4.366***	***17.105***
GQ 10	.728	.445	-1.026	-.951	LQ 10	.717	.451	-.965	-1.071
GQ 11	.603	.490	-.422	-1.826	LQ 11	.646	.478	-.613	-1.628
GQ 12	.602	.490	-.417	-1.830	LQ 12	.546	.498	-.183	-1.971
GQ 13	.960	.195	***-4.732***	***20.443***	LQ 13	.694	.461	-.845	-1.290
GQ 14	.494	.500	.024	-2.004	LQ 14	.682	.466	-.784	-1.388
GQ 15	.633	.482	-.553	-1.698	LQ 15	.811	.392	-1.588	.523
GQ 16	.517	.500	-.067	-2.000	LQ 16	.812	.391	-1.598	.555
GQ 17	.877	.329	-2.293	***3.265***	LQ 17	.498	.500	.010	-2.005
GQ 18	.535	.499	-.140	-1.985	LQ 18	.630	.483	-.537	-1.715
GQ 19	.920	.272	***-3.093***	***7.588***	LQ 19	.582	.494	-.331	-1.895
GQ 20	.612	.488	-.458	-1.794	LQ 20	.664	.473	-.697	-1.518
GQ 21	.821	.383	-1.681	.827	LQ 21	.820	.384	-1.670	.791
GQ 22	.354	.478	.613	-1.628	LQ 22	.655	.476	-.652	-1.579
GQ 23	.910	.288	-2.847	***6.118***	LQ 23	.442	.497	.232	-1.951
GQ 24	.517	.500	-.067	-2.000	LQ 24	.783	.412	-1.375	-.109
GQ 25	.863	.344	-2.119	2.496	LQ 25	.584	.493	-.341	-1.888
GQ 26	.652	.477	-.641	-1.593	LQ 26	.849	.358	-1.952	1.815
GQ 27	.795	.404	-1.464	.143	LQ 27	.685	.465	-.796	-1.369
GQ 28	.658	.475	-.669	-1.557	ELKT	38.248	9.282	-.321	-.696
GQ 29	.699	.459	-.869	-1.247	Reliability Estimates				.889

Note. N=834. The full mark was 56. ELKT represents the English language knowledge test.

The means for individual items varied from .354 to .960 (see Table 1). This indicated a wide range of item-difficulty levels. The standard deviations for individual items ranged from .195 to .500. Six items had a skewness and a kurtosis beyond $\pm$3.000

(GQ2, GQ13, GQ19, LQ1, LQ4 and LQ9), while two items had a kurtosis beyond $\pm$3.000 (GQ17 and GQ23), both of which suggested the non-normal distribution for these items. Despite this, I retained these items at this stage for the following two reasons. Firstly, some of these items would be deleted in exploratory factor analysis (GQ2, GQ13, GQ17, GQ19 and GQ23), and their non-normal distribution would not influence the structural equation modeling (SEM) analysis. Secondly, in the current study, the non-normal distribution at the item level affected the SEM analysis little, since it was a subgroup of test items, rather than a test item, which functioned as an observed variable. The non-normal distribution might disappear after these items combined with other items to form an observed variable.

With reference to the overall test, both values for skewness and kurtosis of the whole test did not exceed the acceptable range, suggesting that the scores of the entire English language knowledge test were normally distributed. In addition, the mean for the entire test corresponded to 38.248, revealing a moderate difficulty level of this test. The standard deviation was 9.282, which indicated moderate individual differences. Moreover, the reliability estimate for internal consistency (Cronbach's alpha) corresponded to .889, suggesting that this English language knowledge test functioned as a reliable measuring instrument. Then, I conducted EFAs on this 56-item English language knowledge test to extract the components (constructs) underlying this test.

III. Exploratory factor analysis for the English language knowledge test

Prior to performing exploratory factor analyses (EFAs) on the English language knowledge test, I examined a matrix of product-moment correlations among test items, the Bartlett's test of sphericity, and the Kaiser-Mayer-Olkin measure of sampling adequacy to see the appropriateness for submitting this measure to EFAs. The result indicated that item correlations were satisfactory. Further, the Bartlett's test of sphericity reached statistical significance ($p < .050$) and the Kaiser-Mayer-Olkin measure of sampling adequacy exceeded .700 (i.e., .841), suggesting that it was suitable to conduct factor analysis on the data collected from the English language knowledge test.

Then, I carried out an array of exploratory factor analyses (EFAs) to extract common components (i.e., constructs) underlying this English language knowledge test. After a number of EFAs were run, the two-component oblique solution (for correlated components) maximized parsimony and interpretability, predicated on (a) the eigenvalues[2] which should be greater than 1.000, (b) the information shown on the scree plot[3], (c) the interpretability of the result, and (d) the inter-component correlation coefficient[4] being .473. Two components with eigenvalues 8.407 and 1.859 respectively were extracted (see Appendix 5 for details). I excluded eighteen items – five lexical items and thirteen grammatical items – due to their low factor loadings (i.e., being below .300)[5] and un-interpretability. Further, I also deleted LQ12 on account of its reduction of the reliability of the first subscale (i.e., Component 1). Finally, thirty-seven items were retained: twenty-one items in the vocabulary subtest and sixteen items in the grammatical subtest.

Test items grouped in Component 1 were constructed to measure students' lexical knowledge. Students were required to access their lexical knowledge base to search for the definitions for assessed words. Or they were required to process a sentence or sentences with a blank embedded in and then get a word from options for the blank to make the entire sentence or sentences meaningful, with their access to their lexical knowledge base and the semantic clues provided by the sentence or sentences. I labeled Component 1 as *lexical knowledge* (LK). On the other hand, test items grouped within Component 2 were designed to assess students' grammatical knowledge. Students were required to process a sentence or sentences with a blank embedded in and then arrive at an appropriate answer for the blank from options to make the overall sentence or sentences grammatically correct and meaningful, with their access to their grammatical knowledge base and the syntactic clues given by the sentence or sentences. I termed Component 2 as *grammatical knowledge* (GK). The English language knowledge test, constituted a vocabulary subtest as well as a grammatical subtest, was intended to gauge

---

[2] In factor analysis with the use of principal components analysis, the eigenvalue of a component denotes that to what extent a component accounts for the total variance. The eigenvalue serves as an index to determine the number of components underlying the collected data to be retained. Traditionally, a component with the eigenvalue greater than 1.000 is retained for further consideration (see Kline, 1994 for further information).

[3] In factor analysis with the use of principal components analysis, the scree plot is another means to decide the number of components underlying the collected data to be retained. The plot is checked to locate a point at which the shape of the curve shifts direction and turns to level off. Components above the break in the plot are retained for further consideration.

[4] If an inter-component correlation coefficient between two components exceeds .500, a consideration is taken that the two components are combined together. If a coefficient is below .200, an orthogonal solution (for uncorrelated components) is preferred.

[5] In the current study, an item with a factor loading of .300 or above was retained after EFAs. This criterion was suggested by Kline (1994). A factor loading of .300 indicates that 9 per cent of the variance is accounted for by the item. This is taken as large enough to suggest that the loading is salient. Further, the criterion is adopted in L2 strategy research where exploratory factor analyses are performed (e.g., Phakiti, 2003; Purpura, 1998a; Vandergrift *et al*., 2006).

participants' lexical knowledge and grammatical knowledge respectively. The result of the EFA indicated that two components (constructs) were extracted underlying the English language knowledge test: LK and GK, which lent support to the presence of the construct validity of this test. The following table shows the reliability estimates for internal consistency for each subscale of test items and for the overall English language knowledge test after EFAs.

Table 2 Reliability estimates for subscales of test items in the English language knowledge test and for the overall English language knowledge test after EFAs

Type of Test Items	Number of Items	Reliability Estimates
LK	21	.829
GK	16	.729
Total	37	.860

Note. LK represents lexical knowledge and GK represents grammatical knowledge.

The reliabilities for LK and GK respectively corresponded to .829 and .729, while the reliability for the total test items was .860 (see Table 2). The result suggested that the contents of the subtests and of the overall test were homogeneous. The vocabulary subtest, the grammatical subtest, as well as the entire English language knowledge test all functioned reliably.

In order to formulate a measurement model of English language knowledge for modeling the relation among EFL students' English language knowledge, strategy use, and their reading test performance, I further categorized test items in each subscale (component) into two subgroups, based on (a) the type of test items, (b) the similar number of test items in each subgroup, and (c) the result of the reliability estimate for each subgroup. More specifically, test items with the same type were basically grouped together; the reliability estimate for each subgroup should be above .500; and items should not reduce the reliability estimate of the subgroup to which they belong. Finally, lexical items measuring LK were divided into LEX1 and LEX2, which functioned as observed variables for LK (a latent variable). LEX1 consisted of 10 items (LQ2, 3, 4, 6, 7, 8, 9, 10, 11 and 13), whilst LEX2 comprised 11 items (LQ1, 14, 15, 16, 20, 21, 22, 24, 25, 26 and 27). Similarly, grammatical items gauging GK were also classified into GRAM1 (GQ1, 3, 4, 6, 7, 8, 9 and 10) and GRAM2 (GQ11, 14, 15, 16, 18, 20, 22 and 24). These two subgroups served as observed variables for GK (a latent variable). Each subgroup was composed of eight test items. The result of the reliability estimates for internal consistency for each subgroup is presented as follows.

Table 3 Reliability estimates for English language knowledge subgroups

Type of Latent Variables	Type of Observed Variables	Number of Items	Reliability Estimates
LK	LEX1	10	.737
	LEX2	11	.703
GK	GRAM1	8	.538
	GRAM2	8	.602

Note. LK represents lexical knowledge and GK represents grammatical knowledge.

As seen in the above table, the reliabilities for these English-language-knowledge subgroups covered from .538 to .737, all of which exceeded the accepted limits ($\alpha > .500$), signifying that the subgroups were reliable to a certain degree. The result also supported the appropriateness for categorizing test items assessing LK into LEX1 and LEX2 and test items measuring GK into GRAM1 and GRAM2.

IV. Distributions and reliabilities of the reading and test-taking strategy use questionnaire

Not distinct from the English language knowledge test, I analyzed the item-level data generating from the reading and test-taking strategy use questionnaire, based on 834 third-graders from six senior high schools. Table 4 provides the descriptive statistics for individual reading and test-taking strategy items and for the overall strategy use questionnaire.

Table 4 Distributions for strategy items in the reading and test-taking strategy use questionnaire and for the entire questionnaire and the reliability estimate for the entire questionnaire

Variable	Mean	Std Dev	Skewness	Kurtosis	Variable	Mean	Std Dev	Skewness	Kurtosis
Item 1	3.868	***.925***	-.920	1.574	Item 38	2.020	1.249	.269	-.897
Item 2	3.679	1.077	-.994	1.237	Item 39	2.144	1.339	.163	-.983
Item 3	3.568	1.299	-.833	.104	Item 40	2.166	1.273	.104	-.948
Item 4	3.384	1.217	-.734	.096	Item 41	2.038	1.323	.278	-.911
Item 5	3.472	1.240	-.839	.277	Item 42	2.959	1.395	-.481	-.645
Item 6	4.440	***.795***	-1.941	***5.847***	Item 43	2.466	1.401	-.062	-1.079
Item 7	2.951	1.479	-.391	-.826	Item 44	3.065	1.332	-.519	-.582
Item 8	3.384	1.242	-.709	.101	***Item 45***	3.872	***.855***	-.986	2.521
Item 9	4.134	***.874***	-1.095	1.675	***Item 46***	3.905	***.898***	-1.010	1.999
Item 10	3.824	1.123	-1.096	1.086	***Item 47***	4.168	***.972***	-1.529	2.859
Item 11	4.058	***.877***	-1.042	1.956	***Item 48***	3.972	***.988***	-1.234	2.060
Item 12	2.964	1.410	-.426	-.765	Item 49	3.802	1.146	-1.147	1.165
Item 13	3.576	1.093	-.807	.468	***Item 50***	4.216	***.966***	-1.565	2.891
Item 14	2.633	1.402	-.154	-.983	***Item 51***	4.299	***.838***	-1.357	2.682
Item 15	3.655	1.063	-.974	1.084	Item 52	3.272	1.243	-.651	-.194
Item 16	2.823	1.313	-.378	-.641	Item 53	2.982	1.272	-.469	-.500
Item 17	3.261	1.132	-.785	.290	Item 54	3.083	1.318	-.509	-.531
Item 18	2.752	1.317	-.278	-.827	***Item 55***	4.058	***.832***	-1.072	2.314
Item 19	3.109	1.200	-.508	-.314	Item 56	3.686	1.078	-.936	.903
Item 20	4.145	***.894***	-1.300	2.473	Item 57	2.585	1.550	-.002	-1.180
Item 21	3.211	1.267	-.632	-.235	Item 58	4.007	1.176	-1.386	1.644
Item 22	3.367	1.148	-.813	.412	***Item 59***	4.231	***.805***	-1.314	***3.234***
Item 23	2.806	1.305	-.313	-.826	***Item 60***	4.106	***.822***	-1.056	2.066
Item 24	2.755	1.438	-.229	-.931	Item 61	3.624	1.117	-.974	.851
Item 25	3.613	1.024	-.978	1.451	Item 62	3.519	1.048	-.814	.727
Item 26	2.892	1.405	-.331	-.867	Item 63	3.839	1.043	-1.170	1.639
Item 27	3.333	1.278	-.811	.021	Item 64	2.294	1.372	.061	-1.029
Item 28	3.330	1.187	-.680	-.004	Item 65	3.646	1.139	-.913	.911
Item 29	2.486	1.302	-.067	-.919	***Item 66***	4.173	***.810***	-1.152	2.498
Item 30	3.639	1.074	-.999	.970	Item 67	2.895	1.217	-.555	-.053
Item 31	3.585	1.050	-.837	.749	Item 68	3.935	1.085	-1.245	1.539
Item 32	3.801	1.025	-1.043	1.412	Item 69	3.922	1.021	-1.194	1.850
Item 33	3.974	***.891***	-.928	1.503	***Item 70***	4.356	***.891***	-1.822	***4.312***
Item 34	2.987	1.203	-.444	-.483	Item 71	3.980	1.067	-1.278	1.734
Item 35	3.034	1.376	-.432	-.756	Item 72	3.121	1.451	-.592	-.401
Item 36	3.525	1.116	-.839	.588	RTSUQ	245.672	34.615	-.147	-.059
Item 37	3.248	1.496	-.533	-.876	Reliability Estimates				.931

Note. N=834. RTSUQ represents the reading and test-taking strategy use questionnaire.

The means for individual items ranged from 2.020 to 4.440, indicating that these strategy items functioned to a certain extent to measure students' strategy deployment (see Table 4). The standard deviations for individual items ranged from .795 to 1.550. The standard deviations of seventeen items were below the 1.000. This suggested that these items had limited variability. I retained these items in this phase, since they may be removed in exploratory factor analysis[6]. If not, later they were to be grouped with other items to form a strategy subgroup and an effect of the limited variability on the following analysis would be minimized. Except three items (i.e., Items 6, 59 and 70)[7] with a kurtosis beyond ±3.000, all values for skewness and kurtosis were within the acceptable limits, suggesting that these items overall were normally distributed.

With respect to the entire questionnaire, both values for skewness and kurtosis were within ±3.000, which indicated that the data of the questionnaire was distributed normally. Additionally, the reliability estimate for internal consistency (Cronbach's alpha) corresponded to .931, demonstrating that this reading and test-taking strategy use questionnaire functioned as a reliable measure. Then, I conducted a series of EFAs on the 72-item questionnaire to summarize the gathered data and to obtain the components (constructs) underlying this scale.

V. Exploratory factor analysis for the reading and test-taking strategy use questionnaire

Similar to the English language knowledge test, I first looked at the data set of the questionnaire to ensure the appropriateness for factor analysis. The procedures adopted to determine the aptness of factor analysis for the English language knowledge test were also applied to the strategy use questionnaire. The result suggested that it was adequate to conduct factor analysis on the data collected from this measure.

Then, in order to extract common components (i.e., constructs) underlying this questionnaire, I carried out a series of exploratory factor analyses (EFAs). After several runs of EFAs, the five-component oblique solution appeared to maximize parsimony and interpretability. However, a close examination of Component 3 showed that there were only three strategy items loading on this component. For the structural equation modeling analysis, these strategy items should be further categorized into at least two strategy subgroups, acting as observed variables for Component 3 (a latent variable). When they

[6] Items 9 and 11 were excluded after exploratory factor analysis.
[7] I remained these three strategy items based on the same reasons detailed in Section II.

were divided into two subgroups, one subgroup would only cover one strategy item, which failed to characterize the subgroup fairly. As a result, I dropped Items 5, 12 and 67 which loaded on Component 3 and re-performed EFAs on the questionnaire data.

After a number of runs of EFAs, the four-component oblique solution maximized parsimony and interpretability on the basis of (a) the eigenvalues being greater than 1.000, (b) the information illustrated on the scree plot, (c) the categories being interpretable, and (d) the inter-component correlation coefficient ranging from .218 to .335. Four components were extracted with eigenvalues 13.473, 3.907, 2.250 and 1.955 respectively (see Appendix 6 for details). I deleted twenty strategy items because their factor loadings were below the cut-off (.300) or they loaded on two components. Moreover, I also dropped Items 17 and 72 since both reduced the reliability of the subscale (component) to which they belonged. Finally, forty-seven strategy items were retained. According to characteristics of strategies within each subscale, I named Component 1 as *monitoring, directing attention and managing the test* (MDAMT), Component 2 as *constructing the meaning and evaluating* (CME), Component 3 as *monitoring and utilizing test questions* (MUTQ), and Component 4 as *evaluating and marking* (EM).

The reading and test-taking strategy use questionnaire was summarized into four components (constructs) after EFAs. In other words, there were four components underlying this measure based on the current data set. In the present study, these four components were defined as strategy use processes in the overall multiple-choice reading comprehension test.

An examination of the *monitoring, directing attention and managing the test* process shows that this process is related to reading and question-answering processes and monitoring plays a substantial and significant role in the entire reading comprehension test. Monitoring functions to check one's comprehension of what has been processed or the current task faced (e.g., *during the reading process, I was aware that I did not understand the meaning of a word*). Then, repeating strategies can be deployed to work on incomprehensible parts (e.g., *when I did not understand the meaning of a sentence, I tried to reread it*). In addition, retrieving-linking strategies or managing-the-test strategies are employed in order to reach a possible answer or optimize test performance (e.g., *when I answered test questions, I tried to recall a part of the passage* or *when I answered test questions, I tried to spend more time on difficult test questions*).

The *constructing the meaning and evaluating* process focuses on the reading process. Local or global reading strategies are manipulated to get a grip on what has been read (e.g., *during the reading process, I tried to use my words to interpret the meaning of the sentence*). In addition, evaluation is present within this strategy use process (e.g., *when I read the passage, I tried to identify the important or less important parts of the passage*).

With respect to the *monitoring and utilizing test questions* process, the question-answering orientation is obvious with the presence of evaluating and monitoring components (e.g., *when I took the test, I tried to use clues from test questions to decide whether to read a particular part of the passage* or *when I read a sentence, I noticed it was related to test questions*).

Finally, within the *evaluating and marking* process, marking strategies with the involvement of assessment are tapped into during the entire reading comprehension test (e.g., *when I read the passage, I tried to mark key points in the passage*).

Based on what is stated above, construct validity is present in this reading and test-taking strategy use questionnaire, given the extracted constructs generally compatible with reading processes involved in Pressley and Afflerbach's (1995) model of constructively responsive reading[8] and the question-answering process where the importance of monitoring is highlighted in Rogers and Bateson's (1991; 1994) model of expert test-takers' test-taking behavior, both of which this questionnaire was grounded on to develop (see Section 3.6.2.2 for details). In addition, as mentioned above, these strategy use processes (i.e., strategy subscales) involve a number of strategies with metacognitive components (e.g., monitoring or evaluation). This is consistent with the notion remarked by Macaro (2004; 2006) and Schraw and Moshman (1995) – strategy groups subsume and are often assessed through a metacognitive strategy or a set of metacognitive strategies. The following table illustrates the reliability estimates for internal consistency for each strategy subscale and for the overall strategy use questionnaire after EFAs.

[8] Pressley and Afflerbach's (1995) model of constructively responsive reading broadly comprises characteristics of (a) the reader response theory in which the transaction between readers and texts is accented; (b) a bottom-up oriented text-processing approach; (c) a top-down oriented text-processing approach; (d) comprehension monitoring processes in which evaluation is often involved; (e) inference-drawing processes.

Table 5 Reliability estimates for strategy subscales and for the overall strategy use questionnaire after EFAs

Type of Processes	Number of Items	Reliability Estimates
MDAMT	20	.871
CME	13	.817
MUTQ	9	.719
EM	5	.783
Total	47	.903

Note. MDAMT represents monitoring, directing attention and managing the test; CME represents constructing the meaning and evaluating; MUTQ represents monitoring and utilizing test questions; EM represents evaluating and marking.

As indicated in Table 5, the strategy use questionnaire consisted of forty-seven reading and test-taking strategy items after EFAs. Each strategy use process (i.e., strategy subscale) encompassed five to twenty strategy items. The MDAMT and the CME processes contained even more strategy items than the EM process, which illustrated a drawback to EFAs in which principal components analysis was adopted. Nonetheless, this shortcoming was excusable, since the aim of EFAs here was to summarize the questionnaire data and pinpoint components (constructs) underlying this scale for the subsequent structural equation modeling analysis. Table 5 also shows that the reliabilities for the subscales ranged from .719 to .871, whereas the reliability for the entire scale was .903. The result suggested that the contents of the subscales and of the overall strategy use questionnaire were homogeneous; fair reliability was present in all the subscales and in the entire questionnaire.

VI. Confirmatory factor analysis for validating the reading and test-taking strategy use questionnaire

In the above EFAs, the four-component solution was obtained for the reading and test-taking strategy use questionnaire. In this section, this four-component solution was further tested by confirmatory factor analysis (CFA) with the use of structural equation modeling (SEM) procedures. In other words, whether the result produced by EFAs still held in CFA was examined. I adopted a model competition procedure for this SEM analysis. More specifically, another two models were constructed grounded on the gathered data: the three-component model and the five-component model. These two models and the four-component model were evaluated simultaneously based on model fit statistics of each model to determine which model described the collected data best.

Prior to performing SEM for CFA, I inspected the z-score of each variable to identify the possible outlier (i.e., the case with the absolute value of the z-score greater than 3.000). Different numbers of cases were located among the three models and they were deleted. In addition, I examined the multivariate normality[9] of the data set. The result of the assessment of multivariate normality indicated that the multivariate kurtosis was within the acceptable limits[10]. Next, I respectively performed CFAs on the three-component, the four-component and the five-component models with the utilization of SEM procedures. The following table presents the model fit indices for these models.

Table 6 The model fit indices for the three-component model, the four-component model and the five-component model

Model fit indices	Levels of acceptable fit	Evaluation results		
		Three-component model	Four-component model	Five-component model
χ^2	Nonsignificant with the p-value > .050	Poor (0.000 with p = unavailable)	Poor (28.005 with p = .000 < .050)	Poor (106.556 with p = .000 < .050)
GFI	> .900	Very good (GFI = 1.000)	Very good (GFI = .983)	Good (GFI = .948)
AGFI	> .900	Poor (AGFI = unavailable)	Good (AGFI = .913)	Poor (AGFI = .843)
CFI	> .950	Very good (CFI = 1.000)	Very good (CFI = .963)	Poor (CFI = .900)
TLI	> .950	Poor (TLI = unavailable)	Poor (TLI = .888)	Poor (TLI = .801)
RMSEA	< .060	Poor (RMSEA = .449)	Poor (RMSEA = .125)	Poor (RMSEA = .157)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation. The three-component model was a just-identified model with zero degree of freedom – the number of distinct sample moments did not depart from the number of distinct parameters to be estimated. Therefore, some model fit indices of which calculation is related to the degrees of freedom were not available.

[9] The assumption should not be violated when the maximum-likelihood estimation procedures are adopted.
[10] As Kline (1998) suggests, when the absolute value of kurtosis is over 10, the distribution of the data is regarded as the non-normal distribution. In the current study, this criterion was adopted.

As shown in Table 6, the three-component model displayed an unsatisfactory goodness-of-fit – the AGFI and the TLI were not available and the RMSEA (.449) was the largest among these three models – although the GFI (1.000) and the CFI (1.000) were larger than the cut-off value. Improvements were found in the four-component model in spite of the lower GFI (.983) and CFI (.963), which was acceptable since both of them exceeded the cut-off value. Moreover, in comparison to those in other two models, the TLI and the RMSEA in the four-component model were better, despite both of them being below the acceptable level. When the five-component model was examined closely, no improvements were observed. The AGFI (.843), the CFI (.900), the TLI (.801) and the RMSEA (.157) did not satisfy the requirements and were worse than those in the four-component model.

To sum up, based on the comparisons of the model fit indices for all the three models, the four-component model described the gathered data more satisfactorily than the three- and the five-component models. Stated simply, four components (i.e., four strategy use processes) represented the collected questionnaire data most appropriately. The following presents the four-component model.

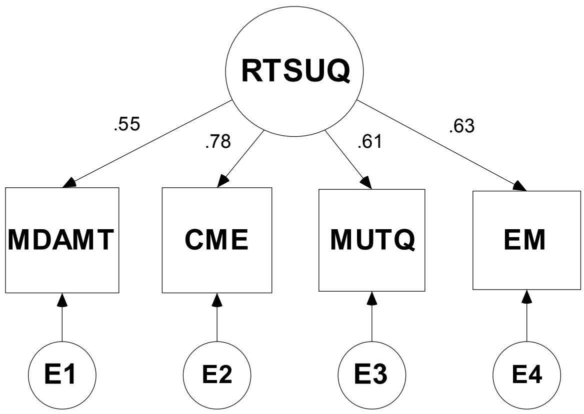


Figure 1 Confirmatory factor analysis of the hypothesized model for the reading and test-taking strategy use questionnaire

The relationships between observed variables and latent variables are profiled by factor loadings which provide information about the extent to which a latent variable can be measured by a given observed variable. It follows that these four components (observed variables at this moment) well represented the reading and test-taking strategy use questionnaire (a latent variable at this moment) in that all the factor loadings, ranging from .553 to .775, exceeded the acceptable level (> .500). The four components (i.e., constructs underlying the reading and test-taking strategy use questionnaire) extracted from exploratory factor analysis (EFA) was supported by the result of confirmatory factor analysis (CFA). To conclude, based on the result of CFA, coupled with that of EFA, the four-component model was accepted as the model of choice.

In order to construct a measurement model of reading and test-taking strategy use for modeling the relation among EFL students' English language knowledge, strategy use, and their reading test performance, I further classified strategy items in each strategy use process into two or five subgroups, based on interpretability and the result of reliability estimates. More specifically, strategy items grouped together generally should share similar attributes with each other and a label can be given. Furthermore, the reliability estimate should exceed .500 and strategy items should not decrease the reliability estimate of the subgroup to which they belong.

Finally, strategy items in the *monitoring, directing attention and managing the test* (MDAMT) process were divided into five strategy subgroups: *monitoring the reading process with negative results* (MRPNP), *repeating* (REP), *monitoring the reading process with positive results* (MRPPR), *retrieving-linking* (RL) and *managing the test with the deployment of test-taking strategies* (MTDTS)[11]. These strategy subgroups functioned as observed variables for the MDAMT process (a latent variable).

Similarly, strategy items included in the *constructing the meaning and evaluating* (CME) process were classified into three strategy subgroups: *constructing the meaning with the deployment of reading strategies* (CMDRS), *evaluating* (EVA) and *interacting with the input* (II)[12]. These three strategy subgroups served as observed variables for the CME process (a latent variable).

Further, strategy items covered by the *monitoring and utilizing test questions* (MUTQ) process were categorized into two subgroups: *monitoring the test-taking process*

[11] The MRPNR subgroup consisted of five items (Items 6, 46, 47, 48 and 66); the REP subgroup three items (Items 20, 50 and 51); the MRPPR subgroup two items (Items 25 and 45); the RL subgroup four items (Items 1, 33, 55 and 59); the MTDTS subgroup six items (Items 56, 60, 63, 68, 69 and 70).
[12] The CMDRS subgroup covered seven items (Items 14, 15, 16, 23, 27, 29 and 42); the EVA subgroup three items (Items 28, 34 and 38); the II subgroup three items (Items 39, 40 and 41).

(MTTP) and *taking advantage of test questions* (TATQ)[13]. Both the MTTP and the TATQ subgroups functioned as observed variables for the MUTQ process (a latent variable).

Additionally, strategy items within the *evaluating and marking* (EM) process were split into two strategy subgroups: *marking incomprehensible parts* (MIP) and *marking key points or options* (MKPO)[14]. These two strategy subgroups served as observed variables for the EM process (a latent variable). The result of the reliability estimates for internal consistency for each subgroup is listed in Table 7.

Table 7 Reliability estimates for strategy subgroups

Type of Latent Variables	Type of Observed Variables	Number of Items	Reliability Estimates
MDAMT	MRPNR	5	.681
	REP	3	.706
	MRPPR	2	.620
	RL	4	.632
	MTDTTS	6	.585
CME	CMDRS	7	.669
	EVA	3	.538
	II	3	.772
MUTQ	MTTP	3	.867
	TATQ	6	.630
EM	MIP	2	.775
	MKPO	3	.653

Note. MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; EM=Evaluating and marking; MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS= Managing the test with the deployment of test-taking strategies; CMDRS=Constructing the meaning with the deployment of reading strategies; EVA=Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP=Marking incomprehensible parts; MKPO=Marking key points or options.

As seen in Table 7, each strategy subgroup consisted of two to seven strategy items. The reliability estimates for these strategy subgroups ranged from .538 to .867, all of which exceeded the accepted level (α > .500), suggesting that these subscales functioned reliably. The result also lent support to the appropriateness for categorizing strategy items in the MDAMT process into the MRPNP, the REP, the MRPPR, the RL and the MTDTS subgroups; strategy items within the CME process into the CMDRS, the

[13] The MTTP subgroup subsumed three items (Items 52, 53 and 54), while the TATQ subgroup six items (Items 3, 4, 22, 26, 43, and 57).
[14] The MIP subgroup encompassed two items (Items 7 and 24), whereas the MKPO subgroup three items (Items 35, 36 and 64).

EVA, and the II subgroups; strategy items in the MUTQ process into the MTTP and the TATQ subgroups; strategy items within the EM process into the MIP and the MKPO subgroups.

VII. Distributions and reliabilities of the reading comprehension test

Similar to the English language knowledge test and the reading and test-taking strategy use questionnaire, I analyzed the item-level data gained from the reading comprehension test, based on 834 third-graders from six senior high schools. The descriptive statistics for reading test items and for the overall reading comprehension test are illustrated in the following table.

Table 8 Distributions for test items in the reading comprehension test and for the entire test and the reliability estimate for the entire test

Variable	Mean	Std Dev	Skewness	Kurtosis	Variable	Mean	Std Dev	Skewness	Kurtosis
RQ 1	.524	.500	-.096	-1.996	RQ 11	.347	.476	.646	-1.586
RQ 2	.514	.500	-.058	-2.001	RQ 12	.607	.489	-.438	-1.813
RQ 3	.832	.374	-1.781	1.173	RQ 13	.613	.487	-.464	-1.789
RQ 4	.628	.484	-.532	-1.721	RQ 14	.616	.487	-.479	-1.775
RQ 5	.739	.440	-1.088	-.818	RQ 15	.831	.375	-1.769	1.132
RQ 6	.751	.433	-1.160	-.655	RQ 16	.561	.497	-.247	-1.944
RQ 7	.830	.376	-1.758	-1.092	RQ 17	.384	.487	.479	-1.775
RQ 8	.803	.398	-1.529	.339	RCT	10.721	3.418	.021	-1.084
RQ 9	.335	.472	.703	-1.510	Reliability Estimates				.744
RQ 10	.799	.401	-1.491	.225					

Note. N=834. The full mark was 17. RCT represents the reading comprehension test.

The means for individual items ranged from .335 to .832 (see Table 8), which suggested a wide range of item-difficulty levels. The standard deviations for individual items ranged from .374 to .500. As for the entire reading test, the mean corresponded to 10.721, indicating a moderate difficulty level of this reading comprehension test. The standard deviation was 3.418, showing moderate individual differences. All values for skewness and kurtosis of individual items and of the whole reading test were within the accepted limits (±3.000), suggesting univariate normality for the distribution of these items and of the entire reading test. In addition, the reliability estimate for internal consistency (Cronbach's alpha) for the overall reading test was .744, demonstrating that this reading comprehension test was a reliable instrument. Then, I submitted these seventeen items to an array of EFAs to uncover the underlying components (constructs).

VIII. Exploratory factor analysis for the reading comprehension test

Like the English language knowledge test, I first inspected the data set of the reading comprehension test to assure the suitability for factor analysis. The procedures utilized for the English language knowledge test were applied to the reading comprehension test. The result suggested that it was appropriate to submit the data collected from this test to factor analysis.

I carried out a series of exploratory factor analyses (EFAs) on the reading comprehension test to extract underlying components (i.e., constructs). After several runs of EFAs, the two-component oblique solution maximized parsimony and interpretability, based on (a) the eigenvalues being greater than 1.000, (b) the information revealed on the scree plot, (c) the interpretability of the result, and (d) the inter-component correlation coefficient being .359. Two components with eigenvalues 3.411 and 1.224 respectively were extracted (see Appendix 8 for details). I dropped Item 8 since its factor loading was below the cut-off (.300). Finally, sixteen items were retained and each component consisted of eight items.

I labeled Component 1, subsuming Items 3, 4, 5, 7, 10, 12, 14 and 15, as *explicit questions* (ExQ). With little inference-drawing, students could arrive at an answer directly after getting a general grip on part of the passage or the whole passage. The items of ExQ assessed participants' ability to read reading passages for facts, details or explicit main ideas. On the other hand, I named Component 2, encompassing Items 1, 2, 6, 9, 11, 13, 16 and 17, as *inferential questions* (InQ). Inferential questions required participants to reason the meaning for what had been read, or infer main ideas (implicit) or true statements (implicit) against the passage. The items of InQ gauged participants' ability to read reading passages for implicit main ideas and to infer meanings from reading passages. This ability, as well as that mentioned above, was intended to be measured by this reading comprehension test. It followed that the construct validity of this reading comprehension test was present and supported to some extent. The following table provides the reliability estimates for internal consistency for each subscale of reading test items and for the overall reading comprehension test after EFAs.

Table 9 Reliability estimates for subscales of reading test items and for the overall reading comprehension test after EFAs

Type of Questions	Number of Items	Reliability Estimates
ExQ	8	.625
InQ	8	.605
Total	16	.734

Note. ExQ represents explicit questions, while InQ represents inferential questions.

The reliabilities for ExQ and InQ were .625 and .605 respectively, whilst the reliability for the total test items was .734 (see Table 9), all of which demonstrated that the contents of each subscale and the overall test were homogeneous. Both subscales and this reading comprehension test functioned reliably. For the subsequent structural equation modeling analysis, originally, I treated ExQ and InQ as latent variables and further categorized test items in ExQ and InQ respectively into two subgroups as observed variables. However, the result of the reliability estimate was unsatisfactory ($\alpha < .500$) for three subgroups out of the four. As a result, I did not divide test items respectively in ExQ and InQ into two subgroups. I viewed ExQ and InQ as observed variables for multiple-choice reading comprehension test performance – a latent variable.

IX. Conclusion

This appendix describes the results of descriptive statistics and EFAs for the English language knowledge test, the reading and test-taking strategy use questionnaire, and the multiple-choice reading comprehension test. The results of EFAs for the English language knowledge test and the multiple-choice reading comprehension test, basically, were similar to the results obtained from the pilot study – two components were extracted. The constructs of these two instruments were validated. As for the reading and test-taking strategy use questionnaire, the result from this analysis based on 834 third-graders from six senior high schools slightly differed from that grounded on 283 third- and second-graders from one senior high school in the pilot study. To illustrate, in the pilot study, five components (constructs) represented the measuring instrument well, while here four components (constructs). This implies that it seems challenging to organize strategy items into different groups in a clear-cut manner, given the nature of strategies themselves, users' characteristics and attributes of tasks encountered.

With the results of exploratory factor analyses as a basis, three measurement models were constructed with regard to English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance. In Chapter Four, in order to examine whether observed variables well represent latent variables in the measurement models of English language knowledge and of reading and test-taking strategy use, I will perform confirmatory factor analyses with the use of structural equation modeling procedures to test these two measurement models. Additionally, I will postulate and test a model which profiles the relationship amongst EFL students' English language knowledge, reading and test-taking strategy use, and their reading comprehension test performance, predicated on theoretical underpinnings and previous studies on L2 reading, L1-L2 reading, and reading or test-taking strategies.

Appendix 5

Exploratory Factor Analysis for the English Language Knowledge Test

The total variance explained by eigenvalues greater than 1.000 is presented below.

Table 1 The total variance explained in the principal components analysis for the English language knowledge test

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**8.407**	15.013	15.013	8.407	15.013	15.013
2	**1.859**	3.319	18.332	1.859	3.319	18.332
3	**1.705**	3.045	21.377	1.705	3.045	21.377
4	**1.370**	2.446	23.823	1.370	2.446	23.823
5	**1.320**	2.357	26.180	1.320	2.357	26.180
6	**1.308**	2.335	28.515	1.308	2.335	28.515
7	**1.223**	2.183	30.699	1.223	2.183	30.699
8	**1.209**	2.159	32.858	1.209	2.159	32.858
9	**1.185**	2.116	34.974	1.185	2.116	34.974
10	**1.176**	2.100	37.074	1.176	2.100	37.074
11	**1.134**	2.024	39.098	1.134	2.024	39.098
12	**1.117**	1.995	41.093	1.117	1.995	41.093
13	**1.110**	1.981	43.075	1.110	1.981	43.075
14	**1.089**	1.944	45.019	1.089	1.944	45.019
15	**1.049**	1.874	46.893	1.049	1.874	46.893
16	**1.042**	1.860	48.753	1.042	1.860	48.753
17	**1.022**	1.824	50.577	1.022	1.824	50.577
18	.996	1.778	52.356			
19	.970	1.733	54.088			
20	.967	1.726	55.814			
21	.947	1.692	57.506			
22	.943	1.684	59.191			
23	.896	1.600	60.791			
24	.887	1.584	62.375			
25	.883	1.576	63.951			
26	.854	1.524	65.476			
27	.839	1.499	66.975			
28	.826	1.475	68.450			
29	.805	1.438	69.888			
30	.802	1.432	71.321			
31	.789	1.408	72.729			
32	.764	1.364	74.093			
33	.757	1.351	75.444			
34	.744	1.329	76.773			
35	.741	1.323	78.096			
36	.729	1.302	79.398			
37	.706	1.260	80.658			

38	.703	1.255	81.913			
39	.691	1.234	83.147			
40	.677	1.209	84.356			
41	.652	1.164	85.521			
42	.639	1.141	86.661			
43	.636	1.136	87.798			
44	.620	1.107	88.904			
45	.604	1.079	89.983			
46	.599	1.070	91.054			
47	.590	1.053	92.107			
48	.568	1.014	93.120			
49	.541	.966	94.087			
50	.520	.928	95.015			
51	.504	.900	95.915			
52	.484	.865	96.780			
53	.477	.851	97.631			
54	.465	.831	98.462			
55	.436	.779	99.241			
56	.425	.759	100.000			

Note. Extraction Method: Principal Components Analysis.

The initial component extraction yielded 17 components with eigenvalues greater than 1.000, all of which explained 50.577% of the total variance (see Table 1). These 17 components were taken into further consideration. The scree plot in the principal components analysis for the English language knowledge test is provided as follows.

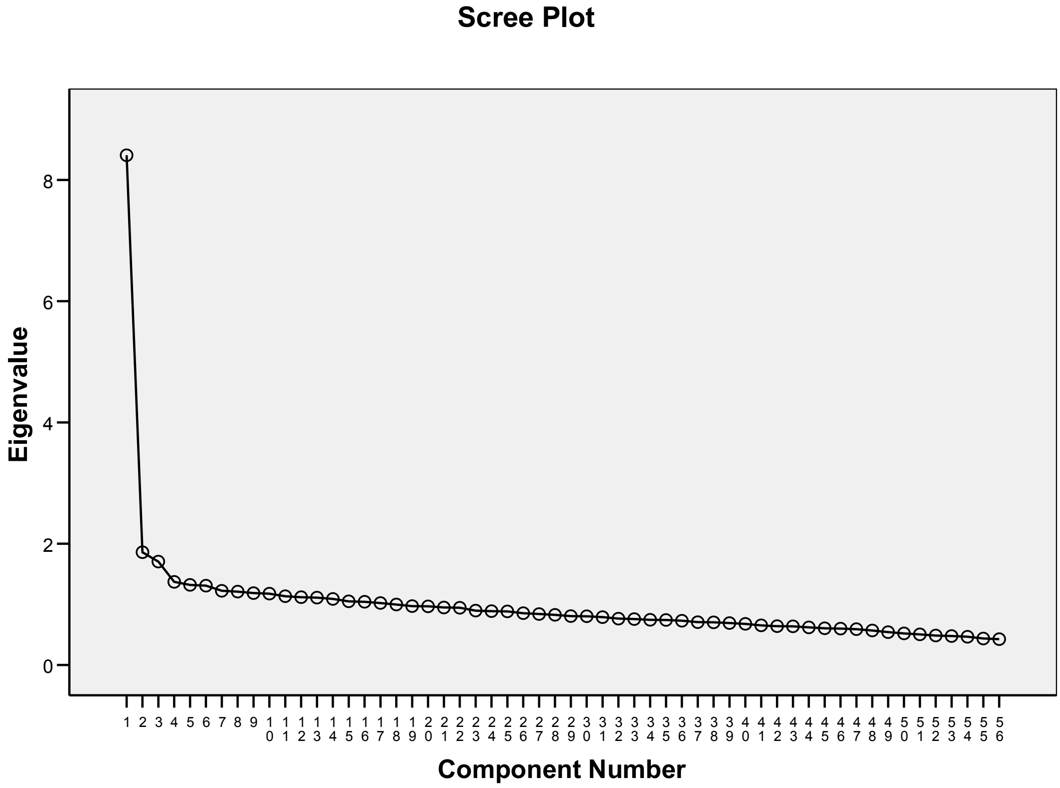


Figure 1 The scree plot in the principal components analysis for the English language knowledge test

As indicated in the above figure, a clear break appeared present after the fourth components, suggesting four components seemed to represent the data best. However, given that this test was originally designed to measure two constructs (grammatical knowledge and lexical knowledge), the two-component solution finally was adopted in order to interpret the data appropriately. A varimax rotation procedure to obtain an orthogonal solution (for uncorrelated components) and a direct oblimin rotation procedure to obtain an oblique solution (for correlated components) were respectively performed. The two-component oblique solution was adopted, since the component correlation matrix revealed that these two components were correlated moderately – the inter-component correlation coefficient was .473 (see Table 2).

Table 2 The component correlation matrix of the English language knowledge test

Component	1	2
1	1.000	.473
2	.473	1.000

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization.

The direct oblimin rotated component pattern matrix is listed as follows.

Table 3 The direct oblimin rotated component pattern matrix(a) of the English language knowledge test

	Component	
	1	2
L24	**.588**	-.111
L7	**.558**	.101
L16	**.549**	.062
L20	**.505**	.042
L26	**.505**	.003
L4	**.502**	-.165
L10	**.459**	.105
L13	**.454**	.136
L11	**.454**	.130
L27	**.451**	.035
L9	**.438**	-.101
L3	**.428**	.152
L6	**.410**	.162
L15	**.398**	.025
L2	**.382**	.175
L21	**.370**	.063
L14	**.364**	.203
G19	***.348***	-.012
L1	**.342**	-.096
L22	**.342**	.127
L12	**.333**	-.097
L8	**.313**	.056
L25	**.308**	.065
L19	***.291***	.184
G27	***.286***	.271
G21	***.271***	.100
L23	***.268***	.239
G13	***.267***	-.046
G23	***.264***	.022
G5	***.260***	.185
G28	***.253***	.243
G2	***.246***	.037
G17	***.245***	.094

G25	***.243***	.135
G12	***.176***	.130
G26	***.156***	-.027
G6	-.034	**.600**
G16	-.054	**.589**
G22	-.144	**.579**
G9	-.092	**.544**
G20	-.061	**.485**
G15	.048	**.410**
G24	.004	**.405**
G14	.191	**.401**
L5	.134	***.390***
G11	-.030	**.386**
G8	.099	**.379**
G18	.097	**.337**
L18	***.141***	***.336***
G7	-.007	**.334**
G1	.046	**.326**
G3	.082	**.325**
G4	.036	**.309**
G10	.061	**.298**
L17	***.097***	***.250***
G29	***.207***	***.240***

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization. a Rotation converged in 8 iterations.

The above table shows that these items were classified into two categories. Then, I followed Crocker and Algina's (1986) suggestions to decide whether test items should be deleted or retained. First of all, a factor loading of an item is primarily placed on only one component. Secondly, an item possesses a high factor loading. In this analysis, an item with a factor loading of .300 or above was retained. Thirdly, an item is included into an appropriate and interpretable category. Finally, an item contributes to the reliability of the subscale. With these suggestions, I deleted L19, L23 and G29, because their factor loadings were below the .300 cut-off. Although the factor loading of G10 was below the cut-off, I still retained this test item, since its factor loading (i.e., .298) approached the cut-off. Given the interpretability, I also dropped G2, G5, G12, G13, G17, G19, G21, G23, G25, G26, G27, G28, L5, L17 and L18. Further, I deleted L12 because it reduced the reliability of the subscale to which it belonged. Finally, thirty-seven test items were retained. The following table reveals the reliability for the entire scale, both subscales (LK and GK – latent variables) and test-item subgroups (observed variables).

Table 4 Reliability estimates for the entire scale, English language knowledge subgroups (latent variables) and test-items subgroups (observed variables)

Type of Latent Variables	Reliability Estimates	Type of Observed Variables	Items Used	Reliability Estimates
LK	.829	LEX1	L2, L3, L4, L6, L7, L8, L9, L10, L11 and L13	.737
		LEX2	L1, L14, L15, L16, L20, L21, L22, L24, L25, L26 and L27	.703
GK	.729	GRAM1	G1, G3, G4, G6, G7, G8, G9 and G10	.538
		GRAM2	G11, G14, G15, G16, G18, G20, G22, and G24	.602
ELKT	.860			

Note. LK represents lexical knowledge and GK represents grammatical knowledge. ELKT=The English language knowledge test.

The reliability for the entire English language knowledge test was .860 (see Table 4), suggesting that the 37-item English language knowledge test was reliable. The reliabilities for lexical knowledge and grammatical knowledge (latent variables) were .829 and .729. These results indicated that both English-language-knowledge subscales were reliable. For the structural equation modeling analysis, test items in both English-language-knowledge subscales were further divided into two test-item subgroups (observed variables). Furthermore, the above table reveals that the reliabilities for these test-item subgroups ranged from .538 to .737. This suggested that these subgroups functioned reliably.

Appendix 6

Exploratory Factor Analysis for the Reading and Test-taking Strategy Use Questionnaire

The total variance explained by components with eigenvalues greater than 1.000 is listed as follows.

Table 1 The total variance explained in the principal components analysis for the reading and test-taking strategy use questionnaire

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**13.703**	19.031	19.031	13.703	19.031	19.031
2	**3.910**	5.431	24.462	3.910	5.431	24.462
3	**2.301**	3.196	27.659	2.301	3.196	27.659
4	**2.251**	3.126	30.785	2.251	3.126	30.785
5	**1.927**	2.676	33.462	1.927	2.676	33.462
6	**1.644**	2.284	35.745	1.644	2.284	35.745
7	**1.605**	2.229	37.975	1.605	2.229	37.975
8	**1.491**	2.071	40.046	1.491	2.071	40.046
9	**1.402**	1.947	41.993	1.402	1.947	41.993
10	**1.346**	1.869	43.862	1.346	1.869	43.862
11	**1.289**	1.790	45.652	1.289	1.790	45.652
12	**1.282**	1.780	47.432	1.282	1.780	47.432
13	**1.203**	1.670	49.102	1.203	1.670	49.102
14	**1.180**	1.639	50.741	1.180	1.639	50.741
15	**1.138**	1.581	52.322	1.138	1.581	52.322
16	**1.080**	1.500	53.822	1.080	1.500	53.822
17	**1.071**	1.487	55.309	1.071	1.487	55.309
18	**1.041**	1.446	56.755	1.041	1.446	56.755
19	**1.006**	1.397	58.152	1.006	1.397	58.152
20	.981	1.363	59.515			
21	.950	1.319	60.834			
22	.905	1.258	62.091			
23	.890	1.237	63.328			
24	.861	1.196	64.524			
25	.850	1.181	65.705			
26	.837	1.162	66.867			
27	.833	1.156	68.024			
28	.814	1.131	69.155			
29	.791	1.098	70.252			
30	.766	1.064	71.317			
31	.741	1.030	72.346			
32	.722	1.003	73.350			
33	.719	.999	74.349			
34	.705	.980	75.329			

35	.679	.943	76.271			
36	.669	.930	77.201			
37	.654	.909	78.110			
38	.645	.896	79.006			
39	.639	.887	79.893			
40	.629	.873	80.766			
41	.618	.858	81.624			
42	.598	.830	82.454			
43	.593	.824	83.278			
44	.573	.796	84.074			
45	.563	.782	84.856			
46	.558	.775	85.631			
47	.544	.756	86.387			
48	.536	.745	87.132			
49	.524	.727	87.859			
50	.514	.714	88.573			
51	.503	.698	89.271			
52	.498	.692	89.964			
53	.481	.669	90.632			
54	.475	.660	91.293			
55	.448	.622	91.914			
56	.437	.607	92.522			
57	.424	.589	93.111			
58	.419	.582	93.693			
59	.405	.562	94.255			
60	.396	.549	94.804			
61	.390	.542	95.346			
62	.375	.521	95.867			
63	.361	.501	96.368			
64	.354	.491	96.859			
65	.338	.470	97.329			
66	.332	.461	97.790			
67	.305	.423	98.213			
68	.302	.420	98.633			
69	.286	.398	99.030			
70	.277	.384	99.415			
71	.226	.314	99.729			
72	.195	.271	100.000			

Note: Extraction Method: Principal Components Analysis.

The initial component extraction yielded 19 components with eigenvalues greater than 1.000, all of which accounted for 58.152% of the total variance (see Table 1). These 19 components were taken into further consideration. The scree plot in the principal components analysis for the reading and test-taking strategy use questionnaire is presented as follows.

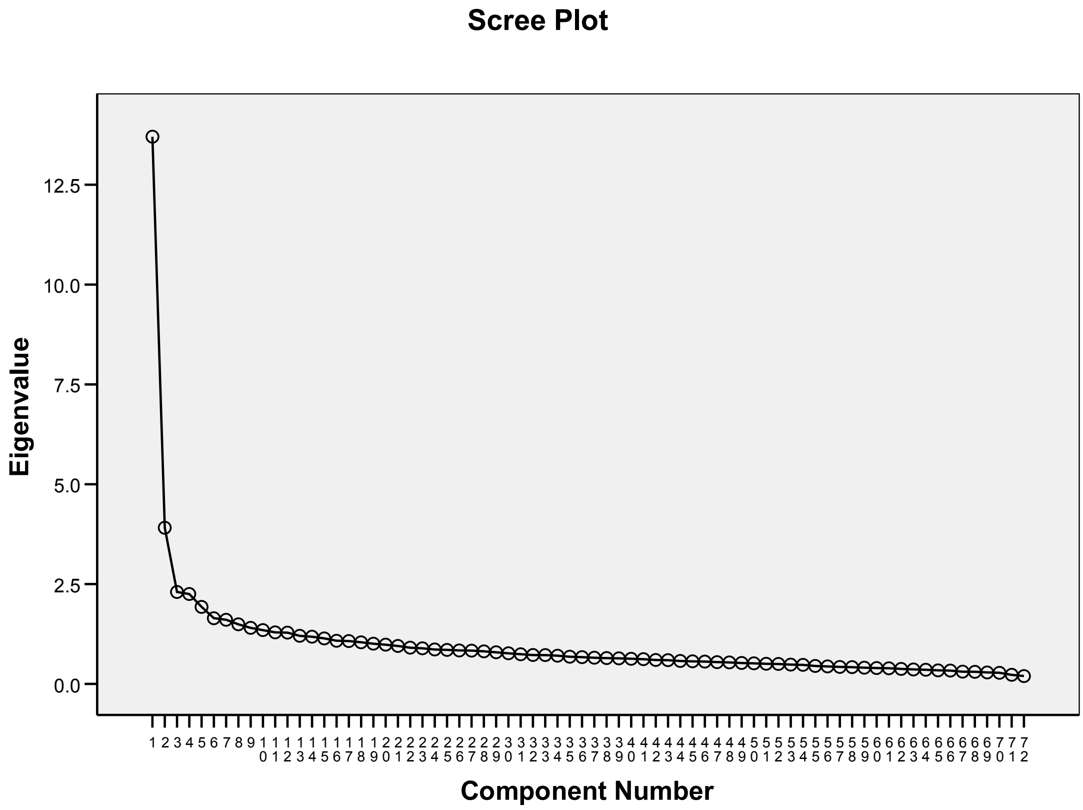


Figure 1 The scree plot in the principal components analysis for the reading and test-taking strategy use questionnaire

As seen in the above figure, a clear break was revealed after the fifth components. This suggested that five components represented the data most fairly, although the curve seemed to turn to level off between the third point and the fourth point. The five components accounted for 33.462% of the total variance. A varimax rotation procedure to obtain an orthogonal solution (for uncorrelated components) and a direct oblimin rotation procedure to obtain an oblique solution (for correlated components) were respectively performed. The five-component oblique solution was adopted since these strategy items were expected to correlate with each other and half of the inter-component correlation coefficients exceeded .200, which ranged from .023 to .334 (see Table 5), although four inter-component correlations were low (below .100).

Table 2 The component correlation matrix of the reading and test-taking strategy use questionnaire

Component	1	2	3	4	5
1	1.000	.257	.023	.334	-.266
2	.257	1.000	-.071	.257	-.295
3	.023	-.071	1.000	.051	.050
4	.334	.257	.051	1.000	-.184
5	-.266	-.295	.050	-.184	1.000

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization.

The direct oblimin rotated pattern matrix is listed below.

Table 3 The direct oblimin rotated component pattern matrix(a) of the reading and test-taking strategy use questionnaire

	Component				
	1	2	3	4	5
Item60	**.668**	-.031	-.114	-.007	-.038
Item11	**.647**	.210	-.140	-.078	.049
Item51	**.615**	-.086	.138	-.054	-.098
Item66	**.612**	-.120	.128	.075	.034
Item20	**.592**	.077	.117	-.111	.011
Item9	**.590**	.221	-.203	-.109	-.016
Item55	**.580**	.008	-.031	.115	.034
Item45	**.573**	.108	-.178	.024	.017
Item48	**.565**	-.048	.049	-.020	-.142
Item46	**.557**	.026	.001	.071	.041
Item50	**.545**	-.121	.111	-.030	-.144
Item59	**.499**	-.057	-.004	.201	-.097
Item56	**.482**	.061	-.146	.005	-.162
Item10	**.468**	.198	-.154	-.055	-.044
Item33	**.467**	.227	-.082	.132	.050
Item25	**.462**	.111	***-.323***	-.010	-.098
Item47	**.455**	-.070	.205	-.041	.004
Item31	**.439**	.272	-.079	-.017	-.070
Item6	**.415**	-.092	***.365***	.012	.007
Item63	**.407**	.089	-.040	.173	.051
Item30	**.396**	.294	.033	.025	-.016
Item68	**.383**	-.085	.200	.052	-.131
Item32	**.371**	***.338***	-.058	.013	.045
Item69	**.371**	.073	.291	.124	.130
Item8	**.345**	.224	-.050	-.065	-.186
Item1	**.328**	-.077	-.240	.228	-.188
Item70	**.304**	-.191	.104	.211	-.196
Item49	***.247***	-.086	.210	.156	-.154
Item58	***.246***	-.106	.112	.226	-.224

Item40	-.152	**.708**	.026	.076	-.060
Item39	-.052	**.689**	-.049	-.066	.019
Item41	-.148	**.661**	.049	-.010	-.081
Item38	-.165	**.548**	.037	.081	-.212
Item23	.007	**.498**	-.089	.057	-.151
Item16	.230	**.451**	-.110	-.050	-.109
Item28	.236	**.433**	-.037	.169	.069
Item42	.085	**.412**	.017	-.068	-.016
Item27	.112	**.394**	.042	.196	.107
Item15	.077	**.386**	.272	.124	.035
Item29	.033	**.361**	.043	.125	-.246
Item17	.195	**.343**	.129	-.094	-.253
Item14	.043	**.334**	-.010	.038	-.207
Item62	.273	**.311**	-.010	.187	.219
Item21	-.006	**.307**	.067	.267	-.141
Item44	.120	***.280***	-.030	.099	-.222
Item18	.209	***.253***	.109	.030	-.123
Item5	-.004	.027	**.689**	.029	-.159
Item12	.025	.122	**.623**	-.001	-.161
Item67	.247	.082	**-.371**	.125	-.194
Item3	-.193	.033	.079	**.616**	.009
Item53	.081	-.059	-.294	**.537**	-.253
Item52	.141	-.072	-.273	**.523**	-.220
Item4	-.029	.002	.103	**.523**	.097
Item26	-.012	.002	-.095	**.516**	.084
Item54	.051	.041	-.230	**.503**	-.220
Item43	-.165	.099	.148	**.495**	-.022
Item22	.076	.166	-.155	**.362**	-.186
Item57	-.035	.119	.059	**.323**	-.235
Item61	.246	.028	-.042	***.276***	.098
Item2	.172	-.021	-.040	***.239***	-.132
Item37	.033	.075	.156	***.236***	.095
Item71	.206	.007	.020	***.235***	.112
Item65	.188	.084	.158	***.222***	.047
Item35	-.039	.106	.048	-.012	**-.731**
Item7	.001	.088	.182	-.153	**-.717**
Item24	.006	.147	.083	-.101	**-.688**
Item64	-.171	***.306***	-.041	.090	***-.478***
Item36	.225	.141	-.043	.041	**-.440**
Item72	.185	-.042	-.066	.125	**-.345**
Item13	.129	.170	-.103	.250	**-.296**
Item34	.099	.257	-.169	.212	***-.289***
Item19	.179	.267	-.082	.101	***-.286***

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization. a Rotation converged in 20 iterations.

The above table indicates that these items were classified into five categories. A further examination of Component Three showed that this component covered only three items. For the subsequent structural equation modeling analysis, these items need to be divided into two subgroups. It follows that there will be one subgroup containing merely one item, which can not represent this subgroup well. Hence, a decision was made to drop these three items and this scale was resubmitted to a series of EFAs.

The total variance explained by components with eigenvalues greater than 1.000 is listed as follows – with Items 5, 12 and 67 being dropped.

Table 4 The total variance explained in the principal components analysis for the reading and test-taking strategy use questionnaire with Items 5, 12 and 67 being dropped

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	**13.473**	19.526	19.526	13.473	19.526	19.526
2	**3.907**	5.663	25.189	3.907	5.663	25.189
3	**2.250**	3.262	28.451	2.250	3.262	28.451
4	**1.955**	2.833	31.284	1.955	2.833	31.284
5	**1.819**	2.636	33.920	1.819	2.636	33.920
6	**1.607**	2.329	36.249	1.607	2.329	36.249
7	**1.499**	2.173	38.423	1.499	2.173	38.423
8	**1.461**	2.117	40.540	1.461	2.117	40.540
9	**1.386**	2.009	42.548	1.386	2.009	42.548
10	**1.301**	1.885	44.433	1.301	1.885	44.433
11	**1.268**	1.837	46.270	1.268	1.837	46.270
12	**1.227**	1.778	48.048	1.227	1.778	48.048
13	**1.194**	1.730	49.778	1.194	1.730	49.778
14	**1.136**	1.646	51.424	1.136	1.646	51.424
15	**1.112**	1.611	53.036	1.112	1.611	53.036
16	**1.073**	1.555	54.590	1.073	1.555	54.590
17	**1.043**	1.511	56.102	1.043	1.511	56.102
18	**1.001**	1.451	57.553	1.001	1.451	57.553
19	.978	1.417	58.970			
20	.932	1.351	60.321			
21	.907	1.315	61.635			
22	.887	1.286	62.921			
23	.882	1.279	64.200			
24	.848	1.228	65.428			
25	.834	1.208	66.636			
26	.821	1.190	67.826			
27	.814	1.179	69.006			
28	.770	1.115	70.121			
29	.763	1.106	71.227			
30	.748	1.085	72.311			
31	.724	1.050	73.361			
32	.721	1.045	74.406			
33	.703	1.019	75.425			
34	.684	.992	76.417			

35	.667	.967	77.383			
36	.656	.951	78.334			
37	.648	.939	79.273			
38	.640	.927	80.200			
39	.627	.909	81.109			
40	.607	.879	81.989			
41	.593	.860	82.849			
42	.592	.858	83.707			
43	.568	.824	84.530			
44	.564	.818	85.349			
45	.544	.788	86.137			
46	.537	.778	86.915			
47	.530	.768	87.682			
48	.527	.764	88.446			
49	.503	.729	89.175			
50	.500	.724	89.899			
51	.484	.701	90.600			
52	.473	.685	91.286			
53	.468	.679	91.964			
54	.452	.655	92.619			
55	.424	.615	93.234			
56	.421	.611	93.845			
57	.405	.588	94.432			
58	.397	.576	95.008			
59	.393	.569	95.577			
60	.370	.536	96.113			
61	.363	.526	96.638			
62	.345	.499	97.138			
63	.336	.487	97.625			
64	.325	.471	98.095			
65	.306	.443	98.539			
66	.295	.428	98.967			
67	.286	.414	99.381			
68	.231	.334	99.715			
69	.197	.285	100.000			

Note. Extraction Method: Principal Components Analysis.

The initial component extraction yielded 18 components with eigenvalues greater than 1.000, all of which accounted for 57.553% of the total variance (see Table 4). These 18 components were taken into further consideration. The scree plot in the principal components analysis for the reading and test-taking strategy use questionnaire is presented as follows.

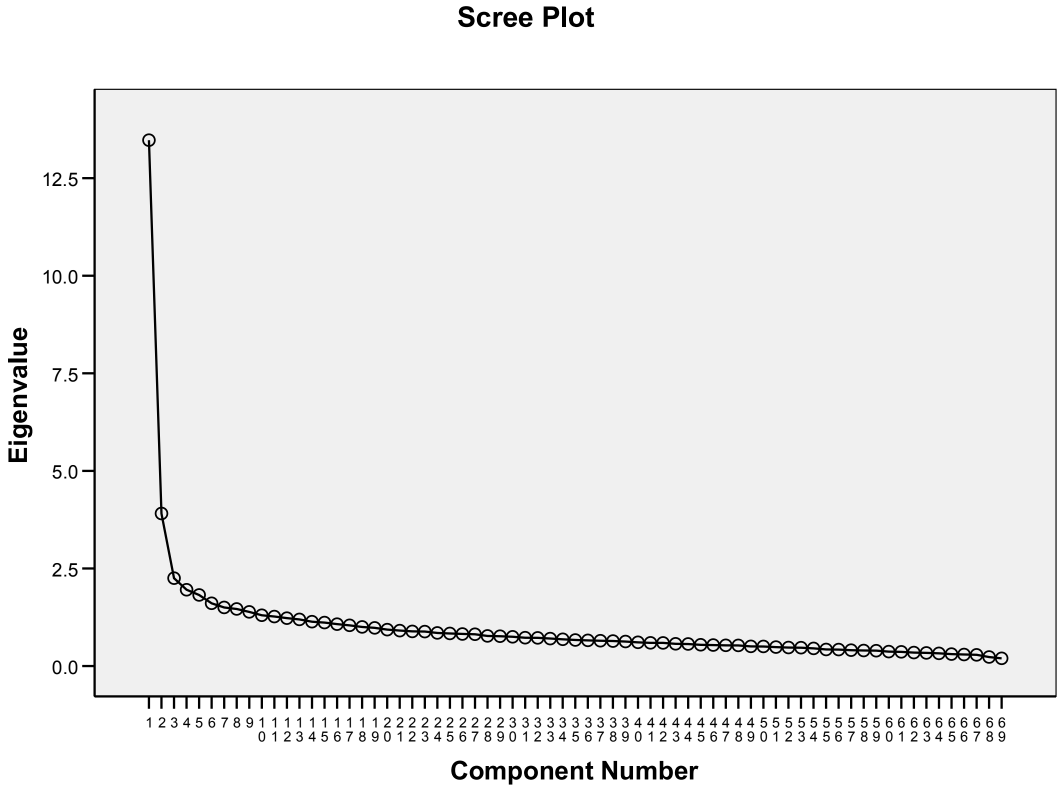


Figure 2 The scree plot in the principal components analysis for the reading and test-taking strategy use questionnaire with Items 5, 12 and 67 being dropped

As displayed in the above figure, the shape of the curve appeared to turn to level off after the fifth component, which indicated that five components represented the data best. However, the result of confirmatory factor analysis with the use of SEM analysis showed that the five-component solution failed to represent the data well (most of the model fit statistics were not satisfactory). Thus, a four-component solution was adopted. The four components accounted for 31.284% of the total variance. A varimax rotation procedure to obtain an orthogonal solution (for uncorrelated components) and a direct oblimin rotation procedure to obtain an oblique solution (for correlated components) were respectively performed. The four-component oblique solution was adopted, since these strategy items were expected to correlate with each other and the component correlation matrix displayed that these four components were correlated moderately – the inter-component correlation coefficients ranged from .218 to .335 (see Table 5).

Table 5 The component correlation matrix of the reading and test-taking strategy use questionnaire – with Items 5, 12 and 67 being dropped

Component	1	2	3	4
1	1.000	.246	.335	.233
2	.246	1.000	.245	.284
3	.335	.245	1.000	.218
4	.233	.284	.218	1.000

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization.

The direct oblimin rotated pattern matrix is presented below.

Table 6 The direct oblimin rotated component pattern matrix(a) of the reading and test-taking strategy use questionnaire with Items 5, 12 and 67 being dropped

	Component			
	1	2	3	4
Item60	***.654***	.054	-.013	.007
Item51	***.630***	-.098	-.060	.120
Item66	***.621***	-.138	.070	-.008
Item11	***.609***	***<u>.336</u>***	-.076	-.114
Item20	***.587***	.075	-.110	.010
Item48	***.572***	-.043	-.025	.157
Item55	***.568***	.083	.113	-.081
Item50	***.564***	-.128	-.035	.159
Item9	***.550***	***<u>.367</u>***	-.109	-.063
Item46	***.546***	.043	.070	-.024
Item45	***.542***	.213	.022	-.061
Item59	***.502***	-.024	.197	.079
Item47	***.471***	-.127	-.041	.046
Item56	***.468***	.149	.002	.119
Item6	***.446***	-.167	.012	.019
Item10	***.437***	***<u>.319</u>***	-.055	-.027
Item33	***.435***	.258	.138	-.072
Item25	***.425***	.239	-.017	.025
Item31	***.410***	***<u>.339</u>***	-.010	.040
Item68	***.407***	-.132	.049	.157
Item63	***.389***	.115	.175	-.050
Item69	***.378***	-.031	.134	-.049
Item30	***.373***	***<u>.309</u>***	.035	.022
Item70	***.335***	-.212	.204	.207
Item8	***.330***	***<u>.307</u>***	-.061	.122
Item1	***.321***	.075	.217	.064
Item49	***.277***	-.142	.156	.174
Item58	***.273***	-.161	.224	.262
Item39	-.107	**.645**	-.037	.060
Item40	-.197	**.635**	.108	.141
Item41	-.186	**.578**	.019	.171
Item23	-.028	**.513**	.075	.158

Item16	.190	**.503**	-.037	.096
Item38	-.188	**.482**	.105	.252
Item28	.195	**.464**	.186	-.080
Item42	.056	**.402**	-.052	.041
Item32	***.335***	**.383**	.024	-.053
Item27	.080	**.381**	.214	-.091
Item14	.030	**.357**	.049	.185
Item29	.024	**.330**	.139	.222
Item34	.082	**.307**	.218	.216
Item15	.069	**.303**	.145	.005
Item62	.234	**.293**	.202	-.171
Item44	.108	**.273**	.108	.248
Item18	.203	**.212**	.040	.157
Item3	-.184	-.027	**.625**	.009
Item4	-.024	-.053	**.529**	-.078
Item53	.076	.100	**.528**	.103
Item26	-.023	.032	**.517**	-.119
Item52	.136	.090	**.514**	.066
Item43	-.155	.002	**.507**	.076
Item54	.043	.152	**.501**	.115
Item22	.063	.208	**.366**	.164
Item57	-.024	.041	**.331**	.234
Item21	-.013	.280	**.282**	.149
Item61	.232	.009	**.279**	-.057
Item13	.124	.235	**.252**	.236
Item37	.035	-.018	**.247**	-.023
Item71	.199	.003	**.238**	-.107
Item2	.176	.000	**.236**	.114
Item65	.189	.037	**.228**	-.026
Item7	.049	.028	-.154	**.739**
Item35	-.001	.081	-.013	**.734**
Item24	.041	.109	-.100	**.707**
Item64	-.168	.245	.098	**.492**
Item36	.233	.175	.041	**.416**
Item72	.199	-.010	.118	**.316**
Item17	.191	.223	-.081	**.311**
Item19	.167	.282	.108	***.298***

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization. a Rotation converged in 25 iterations.

Like the case in the English language knowledge test, I adopted the criteria suggested by Crocker and Algina (1986) to determine to delete or retain strategy items. Furthermore, all strategy items were scrutinized for ambiguity and lack of appropriateness with other items in a subscale. In this analysis, a strategy item with a factor loading .300 or above was accepted. I dropped Items 2, 13, 18, 19, 21, 37, 44, 49, 58, 61, 62, 65 and 71, because their factor loadings were below the .300 cut-off. I also deleted Items 8, 9, 10, 11, 30, 31 and 32 in that they loaded on two components. Further, I excluded Item 17 and

Item 72, since they reduced of the reliability of the subscale to which they belonged. The following table shows the reliability of the entire scale, all subscales and all strategy subgroups.

Table 7 Reliability estimates for the entire scale, all strategy groups (strategy use processes – latent variables) and all strategy subgroups (observed variables)

Type of Latent Variables	Reliability Estimates	Type of Observed Variables	Items Used	Reliability Estimates
MDAMT	.871	MRPNR	6, 46, 47, 48 and 66	.681
		REP	20, 50 and 51	.706
		MRPPR	25 and 45	.620
		RL	1, 33, 55 and 59	.632
		MTDTS	56, 60, 63, 68, 69 and 70	.585
CME	.817	CMDRS	14, 15, 16, 23, 27, 29 and 42	.669
		EVA	28, 34 and 38	.538
		II	39, 40 and 41	.772
MUTQ	.719	MTTP	52, 53 and 54	.867
		TATQ	3, 4, 22, 26, 43 and 57	.630
EM	.783	MIP	7 and 24	.775
		MKPO	35, 36, and 64	.653
RTSUQ	.903			

Note. MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; EM=Evaluating and marking; RTSUQ=The reading and test-taking strategy use questionnaire; MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA=Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP=Marking incomprehensible parts; MKPO=Marking key points or options.

The reliability for the entire strategy use questionnaire was .903, which suggested that the 47-item strategy use questionnaire was reliable. The reliabilities for the four strategy groups (latent variables) were .871, .817, .719 and .783. These results indicated that all strategy groups were reliable. Finally, for the structural equation modeling analysis, strategy items in strategy groups were further divided into several strategy subgroups. The reliabilities for all strategy subgroups (observed variables) ranged from .538 to .867. Such a result suggested that each strategy subgroup functioned reliably.

Appendix 7

A Taxonomy of Reading and Test-taking Strategy Items after Exploratory Factor Analysis

The reading and test-taking strategy use questionnaire consists of four groups with forty-seven strategy items. The four groups were further divided into two, three or five subgroups for the structural equation modeling analysis. The reliabilities of all subgroups were within the accepted limit ($\alpha > .500$), suggesting that these subgroups were reliable (see Table 7 in Appendix 6 for details). The following table presents a taxonomy of reading and test-taking strategy items after exploratory factor analysis.

Table 1 A taxonomy of reading and test-taking strategy items after exploratory factor analysis

I. Monitoring, directing attention and managing the test (MDAMT) – Checking the current task, detecting problems encountered and then deploying retrieving-linking, repeating or managing-the-test strategies in order to solve the problems or perform the test well. (20 items)
(A) Monitoring the reading process with negative results (MRPNR) – Recognizing one's incomprehension, non-concentration, and the difficulty of the text in the course of reading. (5 items)
6. During the reading process, I was aware that I did not understand the meaning of a word.
46. During the reading process, I was aware that I did not understand a part of the passage.
47. During the reading process, I knew that I didn't concentrate.
48. When I read the passage, I was aware of the difficulty of the passage.
66. During the test-taking process, I was aware that I did not understand options.
(B) Repeating(REP) – Attending to one's incomprehensible parts and applying repeating strategies. (3 items)
20. When I did not understand the meaning of a sentence, I tried to reread it.
50. When I did not understand what I read, I tried to read it slowly.
51. When I did not understand the paragraph, I tried to reread it.
(C) Monitoring the reading process with positive results (MRPPR) – Checking and recognizing one's comprehension of what has been processed and sometimes tolerating ambiguity. (2 items)
25. During the reading process, I was aware that I roughly understood the meaning of the sentence although there was a word I did not understand.
45. During the reading process, I was aware that I understood a part of the passage.
(D) Retrieving-linking (RL) – Checking the current task, and retrieving and linking one's cognitive resources or what has been read to deal with the current task. (4 items)

1. When I got the test, I knew what I was going to do first.
33. When I did not understand a part of the passage, I tried to get clues from test questions to help me understand it.
55. When I answered test questions, I tried to recall a part of the passage.
59. When I answered test questions, I tried to find a related paragraph by using clues from test questions.

(E) Managing the test with the deployment test-taking strategies (MTDTS) – Deploying test-taking strategies in order to perform the test well. (6 items)

56. During the question-answering process, I tried to understand the meanings of test questions appropriately.
60. When I answered test questions, I tried to get my answers based on my understanding of the passage.
63. When I answered test questions, I tried to match options with a part of the passage.
68. When I answered test questions, I tried to spend more time on difficult test questions.
69. When I answered test questions, I was ready to change an answer if necessary.
70. I noticed how much time I still had when I took the test.

II. Constructing the meaning and evaluating (CME) – Assessing what has been read and deploying a variety of strategies to process what has been read in order to construct its meaning. (13 items)

(A) Constructing the meaning with the deployment of reading strategies (CMDRS) – Employing local or global reading strategies to construct the meaning of what has been read. (7 items)

14. During the reading process, I tried to substitute a word in the sentence to help me understand the meaning of the sentence.
15. During the reading process, I tried to use my words to interpret the meaning of the sentence.
16. During the reading process, I tried to make an inference about the sentence I read.
23. During the reading process, I tried to associate something else with the sentence I read.
27. When I read the passage, I tried to predict what I was going to read.
29. When I read the passage, I tried to summarize what I read.
42. When I read the passage, I tried to have a picture in mind about what I read.

(B) Evaluating (EVA) – Checking and making a judgement about what is being processed by retrieving and linking what has been processed or one's cognitive resources with what is being processed. (3 items)

28. When I read the passage, I tried to check if my inference was correct.
34. When I read the passage, I tried to identify the important and the less important parts of the passage.
38. When I read the passage, I tried to ask myself questions about what I read.

(C) Interacting with the input (II) – Involving oneself with and responding to what has been read. (3 items)

39. When I read the passage, I tried to relate it to my personal experiences.
40. When I read the passage, I tried to respond to the content of the passage with my personal opinions.
41. When I read the passage, I tried to respond to the content of the passage with my personal feelings.

III. Monitoring and utilizing test questions (MUTQ) – Checking what and how well one has done in the test-taking course; assessing how to approach the current task with the use of test questions in order to perform the test well. (9 items)
(A) Monitoring the test-taking process (MTTP) – Checking what and how well one has done in the test-taking course. (3 items)
52. During the test-taking process, I was aware of what I did. 53. During the test-taking process, I was aware of how was done. 54. During the test-taking process, I was aware of which strategy was used in answering different types of test questions.
(B) Taking advantage of test questions (TATQ) – With test questions as clues, checking and making a judgement about whether what is read is related to test questions, and then deciding how to approach the current task and checking one's strategy deployment. (6 items)
3. When I took the test, I tried to use clues from test questions to decide whether to read a particular part of the passage. 4. When I took the test, I tried to read the passage quickly for particular information. 22. When I read a sentence, I noticed it was related to test questions. 26. During the test-taking process, I read the relevant information about a test question and immediately answered it. 43. When I read the passage, I had test questions in mind. 57. When I answered test questions, I tried to answer them in different orders based on their difficulty.
IV. Evaluating and marking (EM) – Checking and making a judgement about what has been read and conducting marking in order to perform the test well. (5 items)
(A) Marking incomprehensible parts (MIP) – Checking one's comprehension and marking what is not understood. (2 items)
7. When I encountered an unknown word, I tried to mark it. 24. When I read the passage, I tried to mark the sentence that I did not understand.
(B) Marking key points or options (MKPO) – Assessing what has been read and conducting marking in order to perform the test well. (3 items)
35. When I read the passage, I tried to mark key points in the passage. 36. When I read the passage, I tried to remember where key points were in the passage. 64. When I answered test questions, I tried to mark the differences among options.

Appendix 8
Exploratory Factor Analysis for the Multiple-choice Reading Comprehension Test

The total variance explained by components with eigenvalues greater than 1.000 is presented below.

Table 1 The total variance explained in the principal components analysis for the reading comprehension test

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	***3.411***	20.068	20.068	3.411	20.068	20.068
2	***1.224***	7.197	27.265	1.224	7.197	27.265
3	***1.134***	6.673	33.938	1.134	6.673	33.938
4	***1.062***	6.248	40.186	1.062	6.248	40.186
5	***1.002***	5.891	46.077	1.002	5.891	46.077
6	.948	5.577	51.654			
7	.907	5.334	56.987			
8	.882	5.189	62.177			
9	.834	4.908	67.084			
10	.796	4.683	71.767			
11	.784	4.612	76.380			
12	.755	4.443	80.823			
13	.732	4.306	85.129			
14	.659	3.875	89.004			
15	.654	3.849	92.853			
16	.624	3.672	96.525			
17	.591	3.475	100.000			

Note. Extraction Method: Principal Components Analysis.

The initial component extraction yielded five components with eigenvalues greater than 1.000, all of which explained 46.077% of the total variance (see Table 1). These five components were taken into further consideration. The scree plot in the principal components analysis for the reading comprehension test is offered as follows.

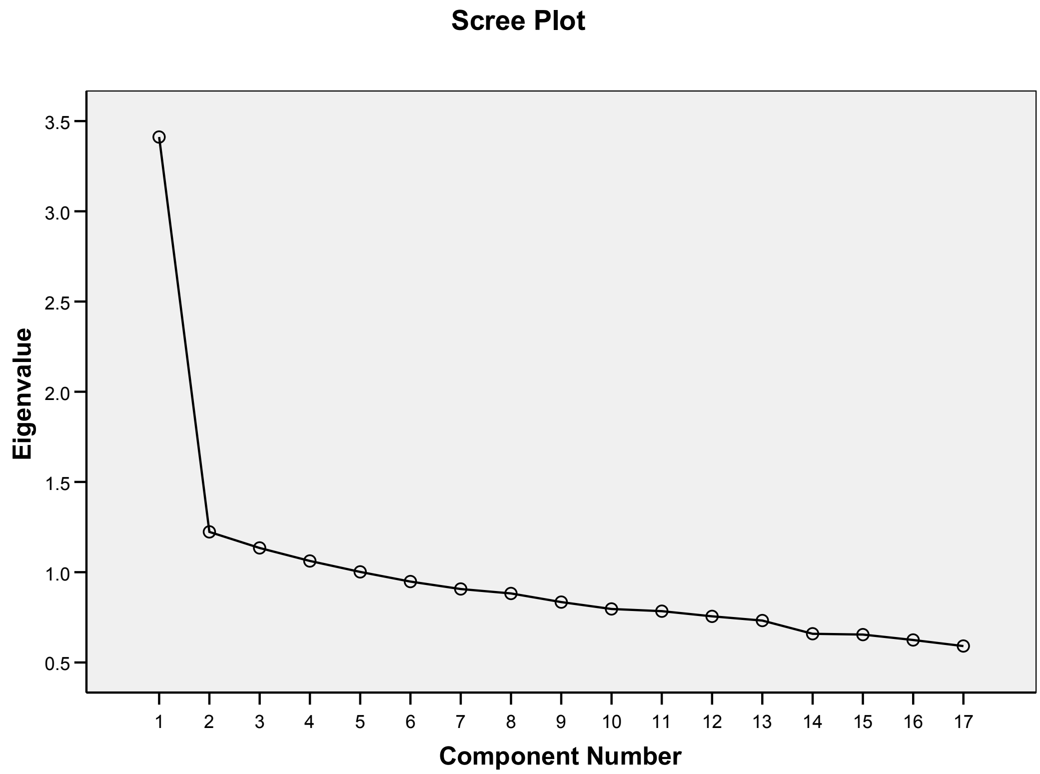


Figure 1 The scree plot in the principal components analysis for the reading comprehension test

As seen in the above figure, the scree plot seemed to level off after the second component, indicating that two components represented the data best. As result, the two-component solution was adopted for the further EFA. A varimax rotation procedure to obtain an orthogonal solution (for uncorrelated components) and a direct oblimin rotation procedure to obtain an oblique solution (for correlated components) were respectively performed. The two-component oblique solution was adopted inasmuch as the component correlation matrix showed that these two components were correlated moderately – the inter-component correlation coefficient was .359 (see Table 2).

Table 2 The component correlation matrix of the reading comprehension test

Component	1	2
1	1.000	.359
2	.359	1.000

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization.

The direct oblimin rotated component pattern matrix is listed in the following table.

Table 3 The direct oblimin rotated component pattern matrix(a) of the reading comprehension test

	Component	
	1	2
R15	**.702**	-.217
R14	**.562**	.077
R10	**.546**	-.015
R12	**.536**	-.007
R7	**.441**	-.052
R4	**.423**	.178
R5	**.381**	.258
R3	**.349**	.067
R8	***.273***	.239
R17	-.230	**.650**
R9	.163	**.571**
R11	-.062	**.510**
R16	.059	**.508**
R1	.015	**.496**
R2	.139	**.357**
R6	.240	**.334**
R13	.236	**.314**

Note. Extraction Method: Principal Components Analysis. Rotation Method: Oblimin with Kaiser Normalization. a Rotation converged in 10 iterations.

The above table indicates that these items were classified into two categories. These test items, next, were examined to see whether they should be dropped. Similar to the case in the English language knowledge test, I adopted the criteria suggested by Crocker and Algina (1986) to determine whether items should be deleted or retained. In this analysis, a test item with a factor loading of .300 or above was accepted. I dropped Item 8 because of its factor loading below the cut-off. Finally, sixteen items were retained. According to factor loadings of test items and interpretability, these items were categorized into two subscales. The following table provides the reliability of the entire scale (a latent variable) and all subscales (observed variables).

Table 4 Reliability estimates for the entire scale (an latent variable), and reading test-items subgroups (observed variables)

Type of Latent Variables	Reliability Estimates	Type of Observed Variables	Items Used	Reliability Estimates
MC RCT	.734	ExQ	Items 3, 4, 5, 7, 10, 12, 14, and 15	.625
		InQ	Items 1, 2, 6, 9, 11, 13, 16 and 17	.605

Note. MC RCT=The multiple-choice reading comprehension test. ExQ=Explicit questions; InQ=Inferential questions.

The reliability for the entire scale was .734 (see Table 4), suggesting that this scale was reliable. The reliabilities for explicit questions and inferential questions (observed variables) were .625 and .605. These results indicated that both reading test-item subscales were reliable.

Appendix 9

Confirmatory Factor Analysis for the Two-component Measurement Model of English Language Knowledge: The Entire Group

The proposed two-component measurement model of English language knowledge was tested by confirmatory factor analysis (CFA) with the use of structural equation modeling (SEM) procedures in which a *confirmatory modeling strategy* was adopted. That is, the hypothesized model was evaluated based on whether this model described the collected data satisfactorily. Prior to performing SEM, I inspected the z-score of each variable to identify the possible outlier (i.e., the case with the absolute value of the z-score greater than 3.000). Three cases were located and they were deleted. In addition, I examined the multivariate normality to ensure that the data set would be generally multivariately distributed and then the parameter estimates estimated by the maximum-likelihood estimation procedures would not be impinged upon. The following table presents the result of the multivariate normality assessment.

Table 1 The assessment of the multivariate normality for the two-component measurement model of English language knowledge for CFA: The entire group

Variable	min	max	skew	c.r.	kurtosis	c.r.
GRAM1	.000	8.000	-.276	-3.249	-.658	-3.870
GRAM2	.000	8.000	.059	.689	-.796	-4.684
LEX1	1.000	10.000	-.598	-7.042	-.614	-3.614
LEX2	2.000	11.000	-.829	-9.756	-.155	-.909
Multivariate					**-.562**	-1.170

Note. N=831.

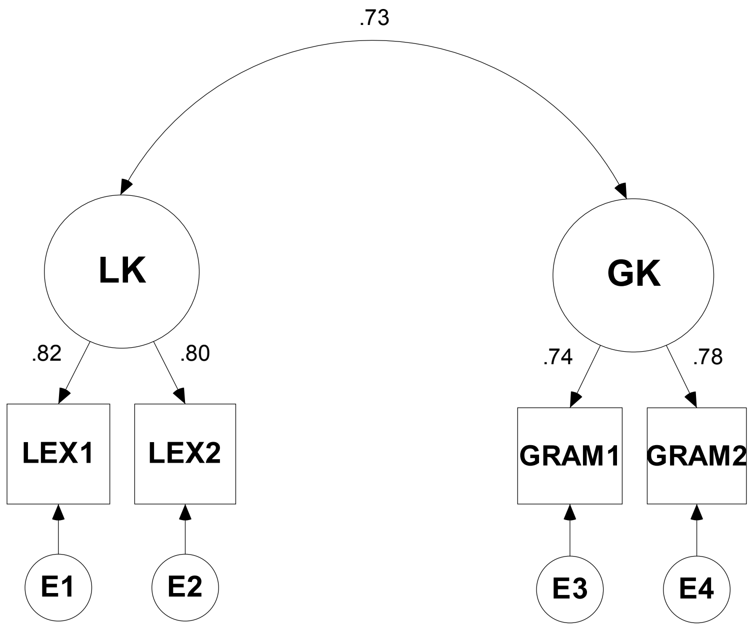
The above table shows that the absolute multivariate kurtosis value was .562, which was within the acceptable limit (< 10). This suggested that the multivariate normality assumption was generally observed. Then, I carried out SEM to test the hypothesized measurement model of English language knowledge. Table 2 presents the summary of the evaluation of model fit.

Table 2 Summary of the evaluation of model fit for the two-component model of English language knowledge: The entire group

Model fit indices	Levels of acceptable fit	Evaluation results
χ^2	Nonsignificant with the p-value > .050	Poor (χ^2 = 3.833, p = .050)
GFI	> .900	**Very good** (GFI = .998)
AGFI	> .900	**Very good** (AGFI = .977)
CFI	> .950	**Very good** (CFI = .997)
TLI	> .950	**Very good** (TLI = .985)
RMSEA	< .060	**Good** (RMSEA = .058)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

As indicated in the above table, the model fit statistics of this accepted model met all the requirements adopted in the current study to determine whether a model was accepted. Thus, it can be concluded that this accepted model, generally, fit the collected data satisfactorily. The accepted measurement model of English language knowledge is shown in the following figure.



LK=Lexical knowledge; GK=Grammatical knowledge. LEX1 consists of ten test items of the vocabulary subtest; LEX2 eleven test items of the vocabulary subtest; GRAM1 eight test items of the grammar subtest; GRAM2 eight test items of the grammar subtest.

Figure 1 The measurement model of English Language Knowledge: The entire group

The factor loadings ranged from .743 to .815 (see Figure 1). The relationships between observed variables and latent variables are profiled by factor loadings which provide information about the extent to which a latent variable can be measured by given observed variables. It follows that these observed variables (LEX1, LEX2, GRAM1 and GRAM2) respectively well explained their latent variables – lexical knowledge (LK) and grammatical knowledge (GK). The result provides evidence for the appropriateness for categorizing test items inclued in LK into LEX1 and LEX2 and test items contained in GK into GRAM1 and GRAM2. The figure also reveals that GK and LK were highly correlated with each other ($r = .729$), implying a close relationship between lexical knowledge and grammatical knowledge as observed in other studies (e.g., Barnett, 1986; Purpura, 1997; 1999). In addition, the correlation coefficient did not approach 1.000, affirming the result generated from EFAs – two components were extracted and suggesting that English language knowledge is well represented by two components (i.e., lexical knowledge and grammatical knowledge).

Appendix 10

Confirmatory Factor Analysis for the Four-component Measurement Model of Reading and Test-taking Strategy Use: The Entire Group

The proposed four-component measurement model of reading and test-taking strategy use was tested by confirmatory factor analysis (CFA) with the use of structural equation modeling (SEM) procedures in which a "model generating strategy" was adopted. In other words, the initial tentative model was proposed and this hypothesized model was evaluated, based on whether this model described the collected data satisfactorily and limited modifications were made if needed. Previous to performing SEM, I examined the z-score of each variable to identify the possible outlier (i.e., the case with the absolute value of the z-score greater than 3.000). Thirty-two cases were located and they were dropped. In addition, I inspected the multivariate normality to ensure that the data set would be generally multivariately distributed and then the parameter estimates estimated by the maximum-likelihood estimation procedures would not be impinged upon. The following table presents the result of the multivariate normality assessment.

Table 1 The assessment of the multivariate normality for the four-component measurement model of reading and test-taking strategy use: The entire group

Variable	min	max	skew	c.r.	kurtosis	c.r.
MKPO	.000	15.000	-.292	-3.376	-.322	-1.859
MIP	.000	10.000	-.334	-3.857	-.727	-4.200
MTDTS	11.000	25.000	-.412	-4.761	-.037	-.214
RL	9.000	20.000	-.353	-4.083	-.196	-1.135
MRPPR	6.000	15.000	-.326	-3.766	-.055	-.321
REP	7.000	15.000	-.727	-8.410	-.034	-.198
MRPNR	12.000	25.000	-.443	-5.118	-.100	-.576
TATQ	8.000	35.000	-.201	-2.322	-.311	-1.799
MTTP	.000	10.000	-.552	-6.386	-.284	-1.639
II	.000	15.000	.184	2.127	-.530	-3.063
EVA	1.000	15.000	-.190	-2.198	-.270	-1.558
CMDRS	7.000	34.000	-.183	-2.113	-.315	-1.819
Multivariate					**18.354**	14.178

Note. N=802.

The value for the multivariate kurtosis was 18.354 (see Table 1), which exceeded the accepted limit (> 10). This suggested the obvious non-multivariate normality of the data. Based on the result of the Mahalanobis-d-squared, I dropped 30 cases of which the value of the Mahalanobis-d-squared exceeded 25. The sample ended up with 772. The following table presents the result of the multivariate normality assessment after 30 cases were dropped.

Table 2 The assessment of the multivariate normality for the four-component measurement model of reading and test-taking strategy use for CFA after 30 cases were dropped: The entire group

Variable	min	max	skew	c.r.	kurtosis	c.r.
MKPO	.000	15.000	-.232	-2.634	-.438	-2.484
MIP	.000	10.000	-.333	-3.778	-.717	-4.069
MTDTS	11.000	25.000	-.369	-4.186	-.054	-.308
RL	9.000	20.000	-.312	-3.535	-.259	-1.468
MRPPR	6.000	15.000	-.286	-3.245	-.059	-.333
REP	7.000	15.000	-.687	-7.798	-.091	-.513
MRPNR	12.000	25.000	-.387	-4.394	-.167	-.948
TATQ	8.000	35.000	-.182	-2.060	-.309	-1.751
MTTP	.000	10.000	-.548	-6.220	-.288	-1.634
II	.000	15.000	.183	2.071	-.535	-3.036
EVA	1.000	15.000	-.192	-2.175	-.303	-1.718
CMDRS	7.000	34.000	-.151	-1.716	-.330	-1.872
Multivariate					**9.657**	7.319

Note. N=772.

As shown in the above table, the absolute multivariate kurtosis value was 9.657, which was within the acceptable limit (< 10). This indicated that the multivariate normality assumption was generally observed. Then, I carried out SEM to test the hypothesized measurement model of reading and test-taking strategy use. At first, the model was postulated as follows. Each component was hypothesized to be related to each other. In addition, the errors of observed variables were postulated to be uncorrelated with each other.

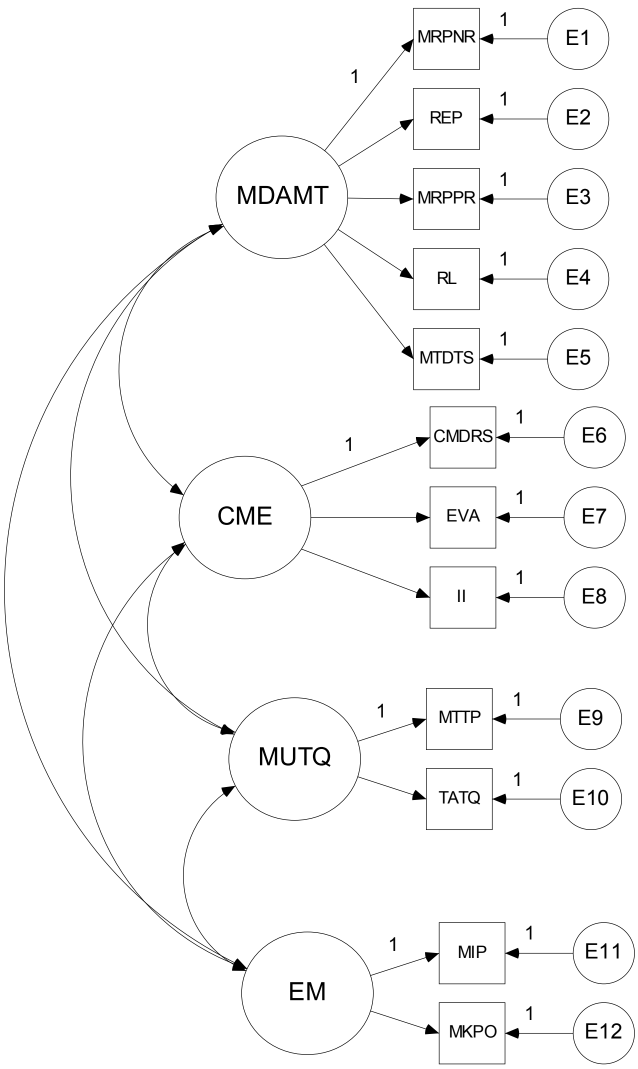


Figure 1 The original hypothesized measurement model of reading and test-taking strategy use: The entire group

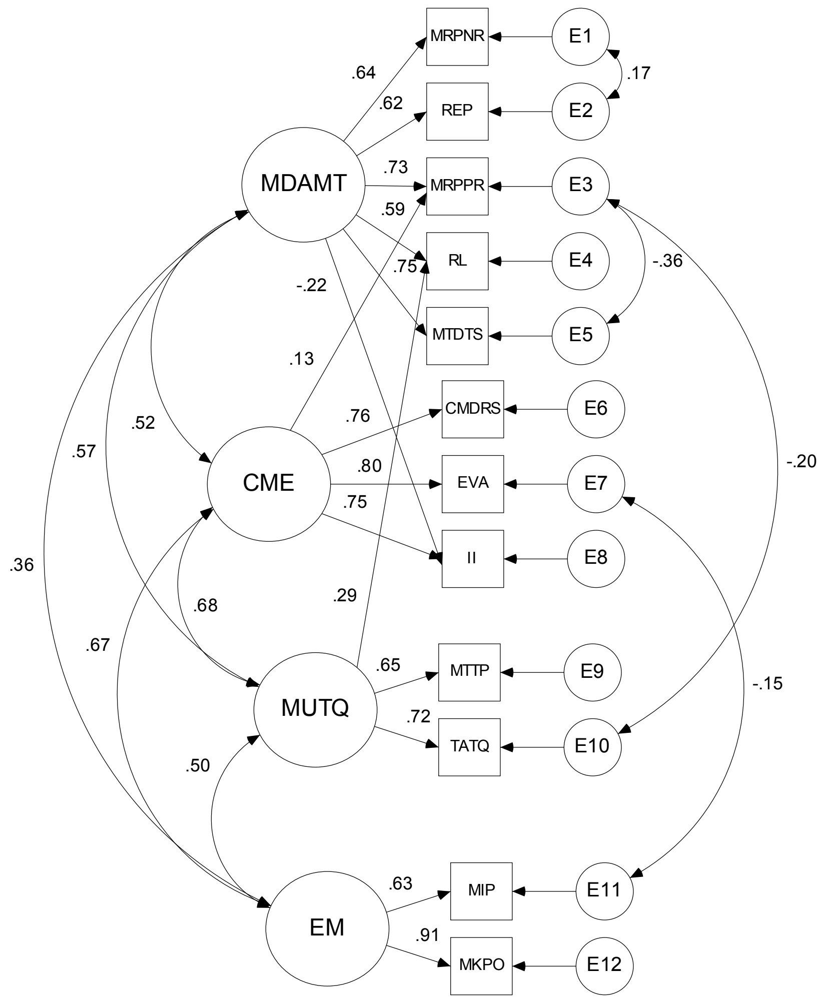
When examining the model fit statistics, I found that the statistics of the GFI (.961), the AGFI (.937), and the CFI (.960) were all above the cut-off value. This suggested that this measurement model seemed to describe the collected data well. However, I noticed that the value of TLI (.945) was below the .950 cut-off, the chi-square statistic was large (χ^2 = 183.083), and the variance estimate of E12 was nonsignificant. Therefore, I made a few adjustments for and respecified the model, grounded on modification indices and interpretability. Finally, a model fitting the sample data well was produced. Table 3 presents the summary of the evaluation of model fit for this accepted model.

Table 3 Summary of the evaluation of model fit for the four-component measurement model of reading and test-taking strategy use: The entire group

Model fit indices	Levels of acceptable fit	Evaluation results
χ^2	Nonsignificant with the p-value > .050	**Good** (χ^2 = 39.471, p = .539)
GFI	> .900	**Very good** (GFI = .991)
AGFI	> .900	**Very good** (AGFI = .984)
CFI	> .950	**Very good** (CFI = 1.000)
TLI	> .950	**Very good** (TLI = 1.001)
RMSEA	< .060	**Very good** (RMSEA = .000)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

The above table indicates the chi-square (χ^2) statistic of 39.471 – much smaller than the previous one. Further, the model fit statistics of this accepted model met all the requirements adopted in the current study to determine whether a model was accepted. This suggested that this accepted four-component measurement model, generally, fit the collected data adequately. The final accepted measurement model of reading and test-taking strategy use is presented in the following figure.



MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ= Monitoring and utilizing test questions; EM=Evaluating and marking; MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA= Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP= Marking incomprehensible parts; MKPO=Marking key points or options.

Figure 2 The final measurement model of reading and test-taking strategy use: The entire group

Reading and test-taking strategy use (RTSU) was represented by four underlying components (latent variables) assessed by two to five observed variables (see Figure 2). The four components subsumed the *monitoring, directing attention and managing the test* process (MDAMT), the *constructing the meaning and evaluating* process (CME), the *monitoring and utilizing test questions* process (MUTQ), and the *evaluating and marking*" process (EM). The factor loadings varied from .593 to .907, except three cross-loadings. The result suggested that these observed variables respectively well represented their latent variables. Furthermore, the result gives evidence for the appropriateness for categorizing strategies covered by the MDAMT process into MRPNR, REP, MRPPR, RL and MTDTS; strategies included in the CME process into CMDRS, EVA and II; strategies contained by the MUTQ process into MTTP and TATQ; strategies subsumed by the EM process into MIP and MKPO, despite the presence of three cross-loadings.

Interesting information is illustrated in the above figure. Three cross-loadings existed in this model. The result offers meaningful information on the nature of these strategies. The *interacting with the input* (II) strategy subgroup loaded on two components (MDAMT and CME), meaning that it measured more than one component. Its cross-loading also revealed that II did not particularly relate to one strategy use component. Intriguingly, II had a high positive relationship with the *constructing the meaning and evaluating* (CME) process, with a factor loading of .751, while it showed a low negative relationship with the *monitoring, directing attention and managing the test* process (MDAMT), with a factor loading of -.222. The result manifests the complex nature of this strategy subgroup. Further, the result indicates that interacting-with-the-input strategies play an important and beneficial role in the CME process; students are willing to spend some time utilizing these strategies to assist in their meaning construction in the reading course, in spite of the fact that the deployment of this type of strategies appears not to directly contribute to building the meaning of the input. However, within the MDAMT process, most strategies are test-taking oriented or monitoring related. It follows that II is less compatible with other strategy subgroups in the MDAMT process and perhaps deploying this strategy subgroup places extra cognitive loads on students in the MDAMT process. Thus, it is reasonable that II is negatively associated with the MDAMT process.

The strategy subgroup *monitoring the reading process with positive results* (MRPPR) respectively displayed a positive relationship with the *monitoring, directing attention and managing the test* (MDAMT) and the *constructing the meaning and evaluating* (CME) processes. The result suggested that MRPPR accounted for not merely the MDAMT process but the CME process as well. Students need this strategy subgroup when (a) they check their comprehension, managing the test or repeating what they do not make sense of and (b) they process the input, constructing its meaning. However, intriguingly, MRPPR carries even more weight in the MDAMT process (.735) than in the CME process (.134). Students should need MRPPR to a certain great degree when they make an effort to build and work out the meaning of what they process. Then, MRPPR should have loaded mainly on the CME process, rather than on the MDAMT process. The reason why MRPPR primarily loaded on the MDAMT process may be that this strategy subgroup often operates with other strategy subgroups (e.g., attention directing or retrieving-linking) in this test-taking context and the MDAMT process happens to contain such strategy subgroups. Or it could be because students' strong test-taking tendencies which lead to the fact that MRPPR loaded more on the MDAMT process covering test-taking strategies to manage the test.

Furthermore, the strategy subgroup *retrieving-linking* (RL) was related to the *monitoring, directing attention and managing the test* (MDAMT) and the *monitoring and utilizing test questions* (MUTQ) processes. The result suggested that RL operated in both the MDAMT and the MUTQ processes. However, it played a more important part in the MDAMT process than in the MUTQ process, since it produced more loadings on the former (.593) than on the latter (.285). The result makes sense, given that retrieving and linking what has been processed with test questions necessitates the deployment of other strategies including in the MDAMT process, such as students' monitoring their current situation and channeling their attention to what is needed in the test-taking course. Additionally, that RL yielded a cross-loading on the MUTQ process is reasonable, because the MUTQ process also covers a strategy subgroup in which test questions are taken advantage of to promote test performance. Utilizing this strategy subgroup appropriately and effectively entails the deployment of retrieving-linking strategies.

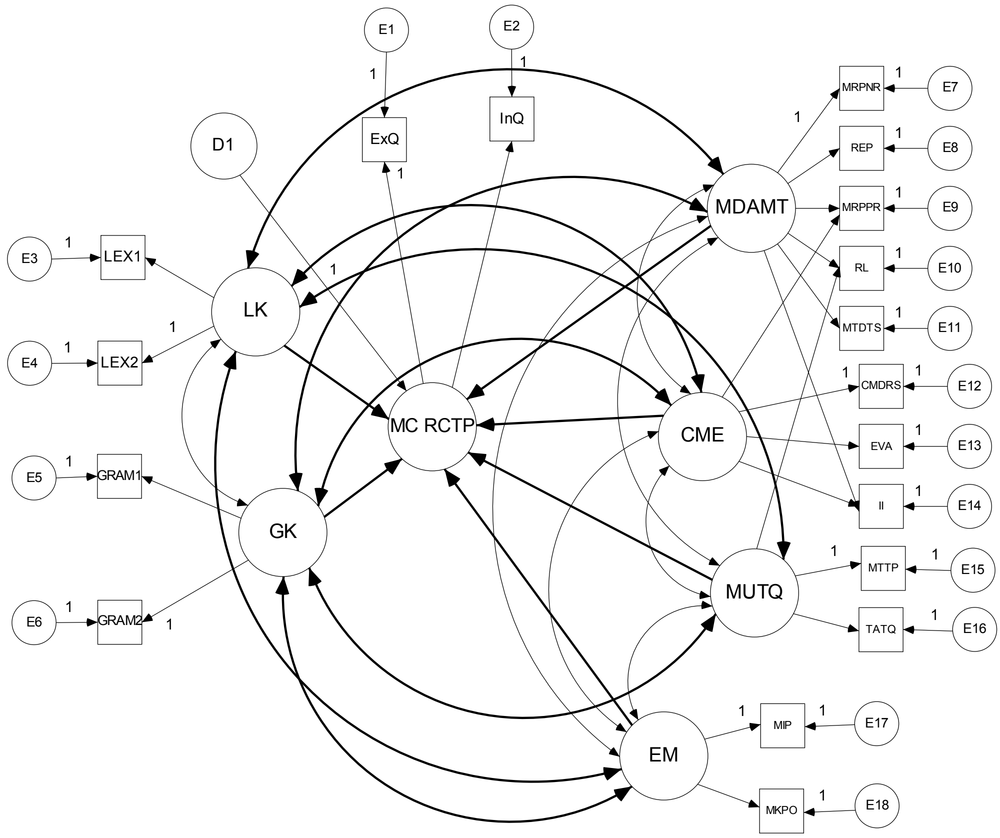
Figure 2 also shows inter-component relationships among components. MDAMT and CME were moderately correlated with each other ($r = .575$). MDAMT was also respectively related to MUTQ and EM ($r = .594$ and .398). Moreover, CME showed a strong relationship with MUTQ and EM respectively ($r = .681$ and .668). Further, MUTQ had a moderate relationship with EM ($r = .505$). These moderate or strong inter-component correlations came as no surprise, since these components were all extracted from the reading and test-taking strategy use questionnaire data by exploratory factor analysis (EFA). In addition, none of these inter-component relationships approached 1.000, lending support to the result of EFA – four components were extracted. This also suggested that reading and test-taking strategy deployment was well characterized by four strategy use processes (i.e., the MDAMT, the CME, the MUTQ, and the EM processes).

Finally, several errors were found to be related to each other significantly. The correlation coefficients ranged from .153 to .361. These error-correlations indicated the presence of some redundant content, measured across strategy subgroups.

Appendix 11

The Full Latent Variable Model regarding the Relationship among English Language Knowledge, Reading and Test-taking Strategy Use, and Multiple-choice Reading Comprehension Test Performance: The Entire Group

The original hypothesized full latent model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance is presented as follows.



"1" functioned as a pre-fixed value for parameter estimation; the initial hypothesized relationships among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance are shown by boldface lines.

Figure 1 The original hypothesized full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance: The entire group

The accepted full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance is presented in the following figure.

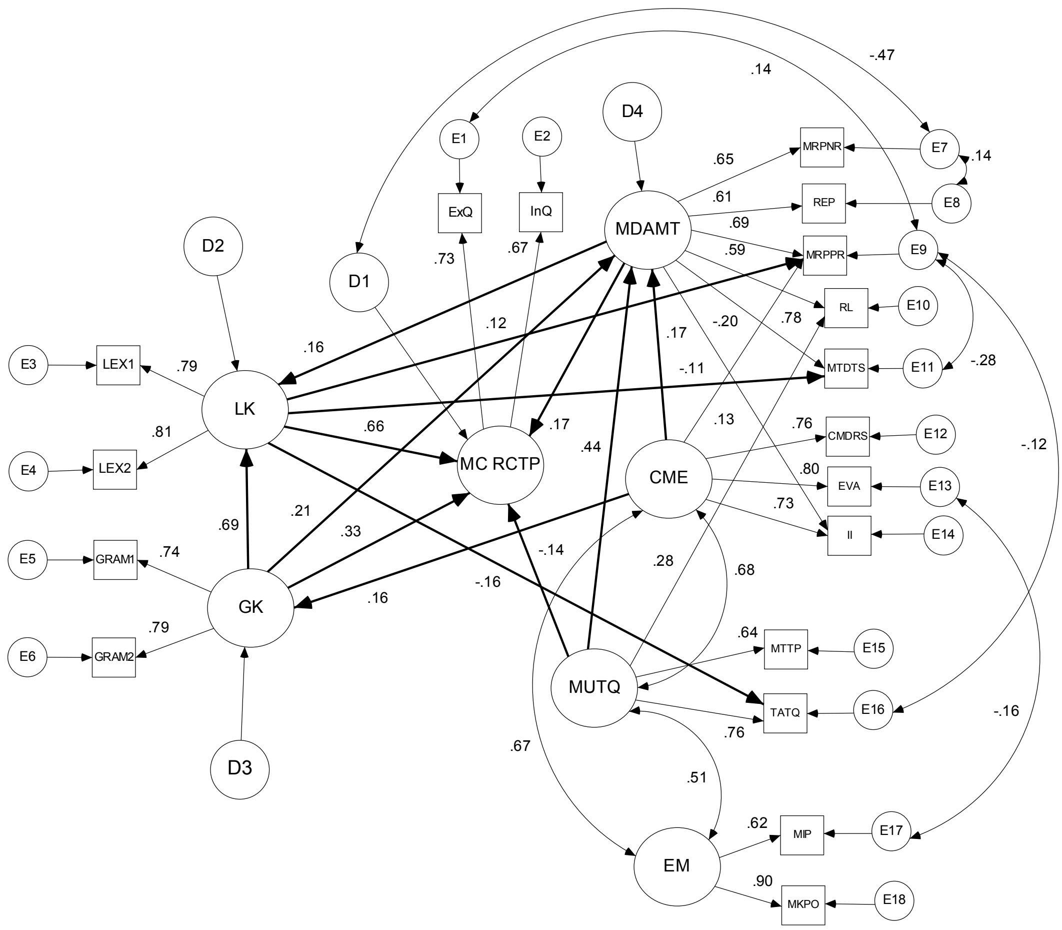


Figure 2 The final full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance: The entire group

The covariance matrix for the full latent variable model regarding the relation among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the overall group is provided as follows.

Table 4 The covariance matrix for the overall group (Part I)

	InQ	ExQ	GRAM1	GRAM2	LEX1	LEX2
InQ	3.978					
ExQ	1.743	3.167				
GRAM1	1.546	1.502	3.340			
GRAM2	1.818	1.766	2.159	4.115		
LEX1	2.299	2.233	1.803	2.120	5.227	
LEX2	2.306	2.240	1.809	2.127	3.314	5.034
MKPO	.393	.382	.374	.440	.610	.612
MIP	.245	.238	.233	.275	.381	.382
TATQ	-.474	-.460	-.239	-.281	-.318	-.319
MTTP	.210	.204	.214	.252	.438	.439
II	.173	.168	.255	.300	.389	.390
EVA	.458	.445	.433	.509	.687	.689
CMDRS	.880	.855	.831	.977	1.319	1.324
MTDTS	.811	.788	.524	.616	.807	.809
RL	.825	.802	.601	.706	1.026	1.029
MRPPR	1.037	1.194	.756	.889	1.283	1.287
REP	.658	.639	.455	.535	.751	.753
MRPNR	.728	.707	.664	.781	1.095	1.099

Table 4 The covariance matrix for the overall group (Part II)

	MKPO	MIP	TATQ	MTTP	II	EVA
MKPO	8.207					
MIP	4.136	6.620				
TATQ	4.857	3.030	26.486			
MTTP	1.924	1.200	5.482	5.254		
II	3.482	2.173	4.609	1.794	10.293	
EVA	3.492	1.684	5.029	2.001	4.072	6.366
CMDRS	6.707	4.184	9.659	3.844	7.821	7.757
MTDTS	1.923	1.200	4.324	1.757	1.110	2.081
RL	2.055	1.282	4.936	2.024	1.481	2.193
MRPPR	1.675	1.045	2.600	1.364	1.217	1.849
REP	1.104	.689	2.324	.998	.641	1.198
MRPNR	1.611	1.005	3.390	1.456	.935	1.747

Table 4 The covariance matrix for the overall group (Part III)

	CMDRS	MTDTS	RL	MRPPR	REP	MRPNR
CMDRS	25.645					
MTDTS	3.997	7.641				
RL	4.212	3.433	4.891			
MRPPR	3.551	2.363	2.500	3.405		
REP	2.301	2.400	1.966	1.715	3.672	
MRPNR	3.356	3.501	2.868	2.503	2.437	7.023

Appendix 12
T-test Analysis of the HEA Group and the LEA Group

The entire participants were divided into two groups: the high English ability (HEA) group and the low English ability (LEA) group. Participants who self-rated their English ability more than 13 (out of 20) were classified into the HEA group, whilst those who self-rated their English ability less than 12 were categorized into the LEA group. The HEA group consisted of 322 participants, while the LEA constituted 512 participants. I performed an independent samples t-test to test whether there were differences between these two groups in English language knowledge, reading and test-taking strategy use, and their reading comprehension test performance. The following table lists the result.

Table 1 The result of t-tests between the HEA group and the LEA group

	Group	N	Mean	Std. Deviation	T value
LEX1	1.00	322	7.708	2.237	6.201**
	2.00	512	6.693	2.340	
LEX2	1.00	322	8.972	2.078	7.022**
	2.00	512	7.881	2.344	
LK	1.00	322	16.680	3.940	7.309**
	2.00	512	14.574	4.223	
GRAM1	1.00	322	5.475	1.776	5.249**
	2.00	512	4.799	1.834	
GRAM2	1.00	322	4.845	1.942	6.745**
	2.00	512	3.898	1.992	
GK	1.00	322	10.320	3.326	6.823**
	2.00	512	8.697	3.355	
ELK	1.00	322	27.000	6.520	8.019**
	2.00	512	23.272	6.548	
MRPNR	1.00	322	21.180	2.681	4.181**
	2.00	512	20.334	3.089	
REP	1.00	322	13.056	1.958	4.401**
	2.00	512	12.410	2.220	
MRPPR	1.00	322	12.236	1.831	7.677**
	2.00	512	11.184	2.071	
RL	1.00	322	16.975	1.987	8.859**
	2.00	512	15.600	2.463	
MTDTS	1.00	322	20.494	2.820	5.598**
	2.00	512	19.264	3.248	
MDAMT	1.00	322	83.941	8.484	7.842**
	2.00	512	78.791	10.313	

CMDRS	1.00	322	22.127	5.212	6.391**
	2.00	512	19.793	5.087	
EVA	1.00	322	8.994	2.460	5.847**
	2.00	512	7.924	2.642	
II	1.00	322	6.776	3.175	3.024**
	2.00	512	6.078	3.291	
CME	1.00	322	37.898	9.020	6.333**
	2.00	512	33.795	9.162	
MTTP	1.00	322	6.935	2.197	6.772**
	2.00	512	5.826	2.365	
TATQ	1.00	322	22.503	5.336	5.153**
	2.00	512	20.617	5.022	
MUTQ	1.00	322	29.438	6.549	6.536**
	2.00	512	26.443	6.373	
MIP	1.00	322	5.966	2.523	2.296*
	2.00	512	5.543	2.693	
MKPO	1.00	322	9.534	2.841	5.319**
	2.00	512	8.424	2.993	
EM	1.00	322	15.500	4.776	4.375**
	2.00	512	13.967	5.019	
RTSU	1.00	322	166.776	21.010	8.661**
	2.00	512	152.996	23.183	
ExQ	1.00	322	6.432	1.611	7.393**
	2.00	512	5.535	1.845	
InQ	1.00	322	4.537	2.034	5.897**
	2.00	512	3.707	1.889	
RCTP	1.00	322	10.969	3.155	7.668**
	2.00	512	9.242	3.173	

Note. **p < .01; *p < .05. LEX1 consists of ten test items of the vocabulary subtest; LEX2 consists of eleven test items of the vocabulary subtest; LK=Lexical knowledge; GRAM1 consists of eight test items of the grammar subtest; GRAM2 consists of eight test items of the grammar subtest; GK=Grammatical knowledge; ELK=English language knowledge. MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA=Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP=Marking incomprehensible parts; MKPO=Marking key points or options. MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ=Monitoring and utilizing test questions; EM=Evaluating and marking; RTSU=Reading and test-taking strategy use. ExQ= Explicit questions; InQ=Inferential questions; RCTP=Reading comprehension test performance.

As shown in the above table, there were significant differences between the HEA group and the LEA group in LEX1, LEX2, LK, GRAM1, GRAM2, GK, ELK, MRPNR, REP, MRPPR, RL, MTDTS, MDAMT, CMDRS, EVA, II, CME, MTTP, TATQ, MUTQ, MIP, MKPO, EM, RTSU, ExQ, InQ and RCTP.

With respect to English language knowledge, the HEA group significantly performed better than the LEA group respectively in the LEX1 subgroup (Mean = 7.708 vs. 6.693), the LEX2 subgroup (Mean = 8.972 vs. 7.881), the LK group (Mean = 16.680 vs. 14.574), the GRAM1 subgroup (Mean = 5.475 vs. 4.799), the GRAM2 subgroup (Mean = 4.845 vs. 3.898), the GK group (Mean = 10.320 vs. 8.697) and the overall test (Mean = 27.000 vs. 23.272) despite the limited mean difference. Such results indicate cross-group discrepancies in English language knowledge.

As for reading and test-taking strategy use, the HEA group significantly had a stronger tendency to their strategy deployment than the LEA group in the following strategy subgroups: MRPNR (M = 21.180 vs. 20.334), REP (Mean = 13.056 vs. 12.410), MRPPR (Mean = 12.236 vs. 11.184), RL (Mean = 16.975 vs. 15.600), MTDTS (Mean = 20.494 vs. 19.264), CMDRS (Mean = 22.127 vs. 19.793), EVA (Mean = 8.994 vs. 7.924), II (Mean = 6.776 vs. 6.078), MTTP (Mean = 6.935 vs. 5.826), TATQ (Mean = 22.503 vs. 20.617), MIP (Mean = 5.966 vs. 5.543), and MKPO (Mean = 9.534 vs. 8.424). Additionally, in comparison with the LEA group, the HEA group also displayed a stronger tendency in four strategy use processes (Mean = 83.941 vs. 78.791 for the MDAMT process; 37.898 vs. 33.795 for the CME process; 29.438 vs. 26.443 for the MUTQ process; 15.500 vs. 13.967 for the EM process) and the entire strategy use (Mean = 166.776 vs. 152.996). The result suggests that the HEA group's strategy employment differs from the LEA group's strategy use.

As far as reading comprehension test performance was concerned, the result was consistent with that in English language knowledge. The HEA group significantly outperformed the LEA group respectively in the ExQ subgroup (Mean = 6.432 vs. 5.535), the InQ subgroup (Mean = 4.537 vs. 3.707) and the entire test (Mean = 10.969 vs. 9.242), although the mean discrepancy was limited. These results suggest variations between the HEA group and the LEA group in their performance on the reading comprehension test. To conclude, the above mentioned results lend support to the subsequent analysis – whether these cross-group differences can be located in the SEM analysis.

Appendix 13
Confirmatory Factor Analysis for the Two-Component Measurement Model of English Language Knowledge: The HEA Group

Based on the data collected from the HEA group, the proposed two-component measurement model of English language knowledge was tested by confirmatory factor analysis (CFA) with the use of structural equation modeling (SEM) procedures in which a *confirmatory modeling strategy* was adopted. Prior to performing SEM, I inspected the z-score of each variable to identify the possible outlier (i.e., the case with the absolute value of the z-score greater than 3.000). No cases were located. In addition, I examined the multivariate normality to ensure that the data set would be generally multivariately distributed and then the parameter estimates estimated by the maximum-likelihood estimation procedures would not be impinged upon. The following table presents the result of the multivariate normality assessment.

Table 1 The assessment of the multivariate normality for the two-component measurement model of English language knowledge for CFA: The HEA group

Variable	min	max	skew	c.r.	kurtosis	c.r.
GRAM1	1.000	8.000	-.394	-2.884	-.656	-2.402
GRAM2	.000	8.000	-.098	-.715	-.760	-2.784
LEX1	1.000	10.000	-1.036	-7.592	.278	1.018
LEX2	2.000	11.000	-1.191	-8.723	.834	3.055
Multivariate					**2.484**	3.217

Note. n=322.

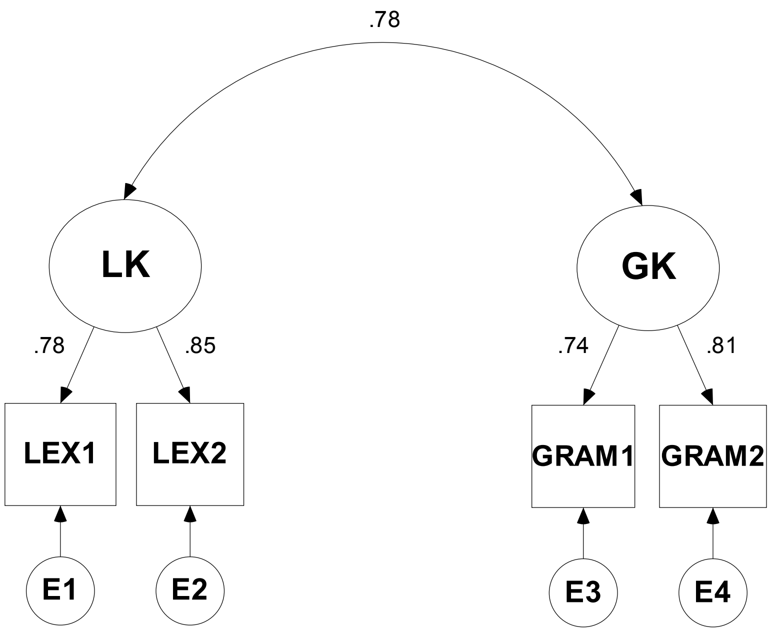
The absolute multivariate kurtosis value was 2.484 (see Table 10), which was within the acceptable limit (< 10). This suggested that the multivariate normality assumption was generally observed. Then, I carried out SEM to test the hypothesized measurement model of English language knowledge for the HEA group. Table 2 presents the summary of the evaluation of model fit.

Table 2 Summary of the evaluation of model fit for the two-component model of English language knowledge: The HEA group

Model fit indices	Levels of acceptable fit	Evaluation results
χ^2	Nonsignificant with the p-value > .050	**Good** (χ^2 = .091, p = .763)
GFI	> .900	**Very good** (GFI = 1.000)
AGFI	> .900	**Very good** (AGFI = .999)
CFI	> .950	**Very good** (CFI = 1.000)
TLI	> .950	**Very good** (TLI = 1.011)
RMSEA	< .060	**Very good** (RMSEA = .000)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

The above table indicates that the model fit statistics of this accepted model satisfied all the requirements adopted in the current study to determine whether a model was accepted. Thus, it can be concluded that this accepted measurement model of English language knowledge satisfactorily described the collected data related to the HEA group. The accepted measurement model of English language knowledge for the HEA group is shown in the following figure.



LK=Lexical knowledge; GK=Grammatical knowledge. LEX1 consists of ten test items of the vocabulary subtest; LEX2 eleven test items of the vocabulary subtest; GRAM1 eight test items of the grammar subtest; GRAM2 eight test items of the grammar subtest.

Figure 1 The measurement model of English language knowledge: The HEA group

As shown in Figure 1, the factor loadings ranged from .740 to .850. This suggested that observed variables (LEX1, LEX2, GRAM1 and GRAM2) respectively well explained their latent variables – lexical knowledge (LK) and grammatical knowledge (GK). Further, the figure also displays that GK and LK were strongly correlated with each other, with a correlation coefficient of .783, not close to 1.000. This indicated that English language knowledge was represented appropriately by two components (i.e., lexical knowledge and grammatical knowledge) based on the HEA group data.

Appendix 14

Confirmatory Factor Analysis for the Four-component Measurement Model of Reading and Test-taking Strategy Use: The HEA Group

Based on data gathered from the HEA group, the proposed four-component measurement model of reading and test-taking strategy use was tested by confirmatory factor analysis (CFA) with the use of structural equation modeling (SEM) procedures in which a *model generating strategy* was adopted. Previous to performing SEM, I examined the z-score of each variable to identify the possible outlier (i.e., the case with the absolute value of the z-score greater than 3.000). Four cases were located and they were deleted. In addition, I inspected the multivariate normality to ensure that the data set would be generally multivariately distributed and then the parameter estimates estimated by the maximum-likelihood estimation procedures would not be impinged upon. The following table presents the result of the multivariate normality assessment.

Table 1 The assessment of the multivariate normality for the four-component measurement model of reading and test-taking strategy use: The HEA group

Variable	min	max	skew	c.r.	kurtosis	c.r.
MKPO	.000	15.000	-.422	-3.070	.114	.416
MIP	.000	10.000	-.397	-2.889	-.569	-2.070
MTDTS	11.000	25.000	-.506	-3.683	-.054	-.197
RL	10.000	20.000	-.484	-3.520	-.070	-.254
MRPPR	6.000	15.000	-.496	-3.608	.423	1.541
REP	7.000	15.000	-.910	-6.624	.573	2.087
MRPNR	12.000	25.000	-.539	-3.926	.091	.332
TATQ	8.000	35.000	-.402	-2.929	-.106	-.385
MTTP	.000	10.000	-.862	-6.276	.309	1.123
II	.000	15.000	.064	.466	-.571	-2.078
EVA	2.000	15.000	-.317	-2.307	-.257	-.936
CMDRS	7.000	34.000	-.465	-3.388	.050	.183
Multivariate					**25.159**	12.238

Note. n=318.

The value for the multivariate kurtosis was 25.159 (see Table 1), which exceeded the acceptable limit (> 10). This suggested the apparent non-multivariate normality of the data. Based on the result of the Mahalanobis-d-squared, I dropped 21 cases of which the value of the Mahalanobis-d-squared exceeded 23. The sample ended up with 297. The following table presents the result of the multivariate normality assessment after 21 cases were dropped.

Table 2 The assessment of the multivariate normality for the four-component measurement model of reading and test-taking strategy use for CFA after 21 cases were dropped: The HEA group

Variable	min	max	skew	c.r.	kurtosis	c.r.
MKPO	2.000	15.000	-.244	-1.713	-.352	-1.239
MIP	.000	10.000	-.382	-2.686	-.531	-1.868
MTDTS	12.000	25.000	-.359	-2.527	-.287	-1.009
RL	12.000	20.000	-.294	-2.070	-.434	-1.525
MRPPR	7.000	15.000	-.276	-1.939	-.046	-.160
REP	7.000	15.000	-.789	-5.549	.284	.999
MRPNR	12.000	25.000	-.411	-2.891	-.194	-.684
TATQ	8.000	35.000	-.345	-2.428	-.066	-.233
MTTP	.000	10.000	-.876	-6.163	.390	1.372
II	.000	15.000	.029	.206	-.556	-1.956
EVA	2.000	15.000	-.336	-2.365	-.277	-.975
CMDRS	7.000	34.000	-.417	-2.937	.058	.204
Multivariate					**8.155**	3.833

Note. n=297.

As shown in the above table, the absolute multivariate kurtosis value was 8.155, which was within the acceptable limit (< 10). This indicated that the multivariate normality assumption was generally observed. Then, I carried out SEM to test the hypothesized measurement model of reading and test-taking strategy use. At first, the model was hypothesized as follows.

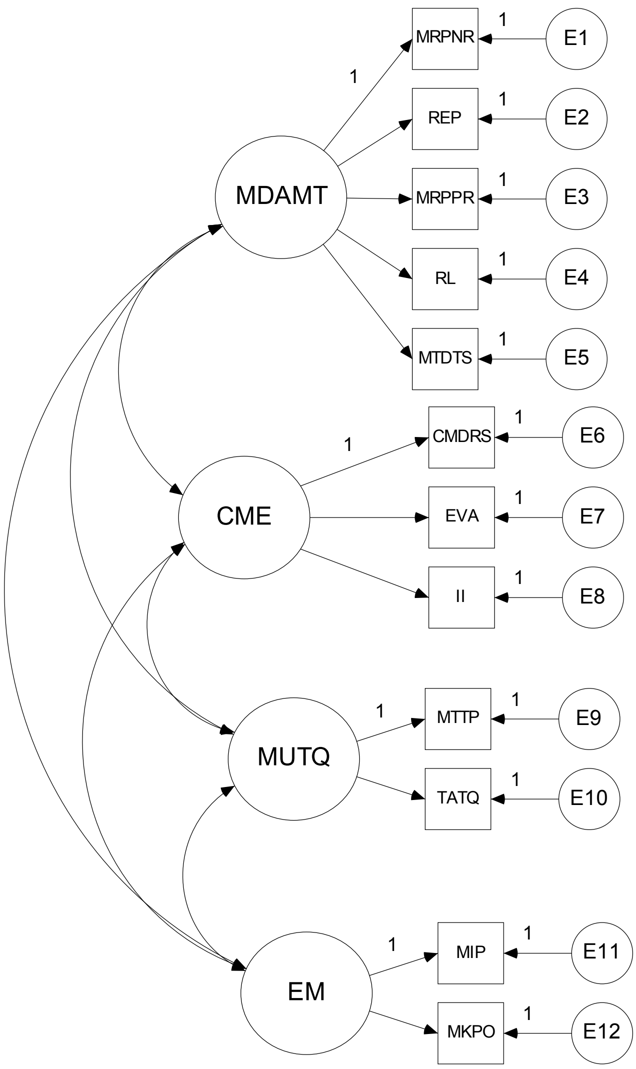


Figure 1 The original hypothesized measurement model of reading and test-taking strategy use: The HEA group

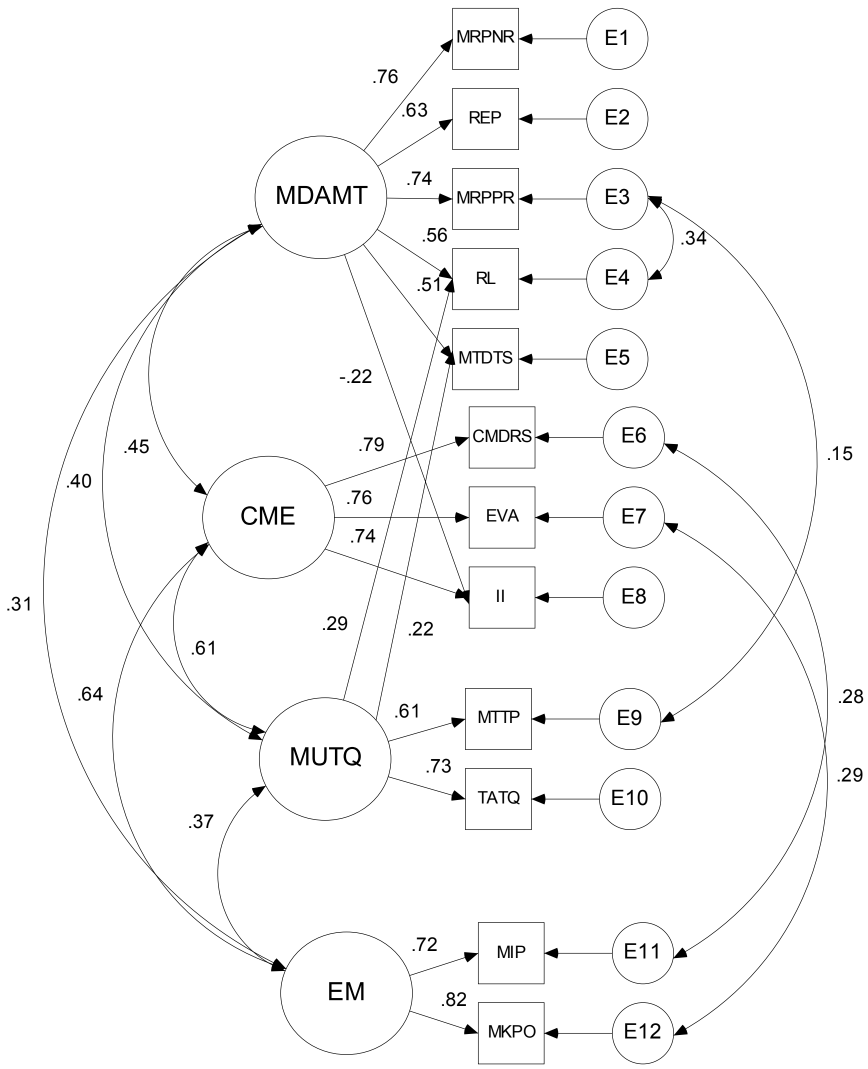
When examining the model fit statistics, I found that the statistics of the GFI (.961) and the AGFI (.901) were both above the .900 cut-off. This implied that this measurement model appeared to describe the collected data well. However, I noticed that the values of the CFI (.942) and the TLI (.920) were below the .950 cut-off, the chi-square statistic (χ^2 = 118.315) was large, and the variance estimate of E12 was nonsignificant. Consequently, I made a few adjustments for and respecified the model, grounded on modification indices, interpretability, and the measurement model of reading and test-taking strategy use for the entire group. Finally, a model fitting the sample data well was produced. Table 3 provides the summary of the evaluation of model fit for this accepted model.

Table 3 Summary of the evaluation of model fit for the four-component model of reading and test-taking strategy use: The HEA group

Model fit indices	Levels of acceptable fit	Evaluation results
χ^2	Nonsignificant with the p-value > .050	**Good** (χ^2 = 44.272, p = .335)
GFI	> .900	**Very good** (GFI = .977)
AGFI	> .900	**Very good** (AGFI = .956)
CFI	> .950	**Very good** (CFI = .997)
TLI	> .950	**Very good** (TLI = .996)
RMSEA	< .060	**Good** (RMSEA = .016)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

The above table reveals the chi-square (χ^2) statistic of 44.272 – even smaller than the previous one. In addition, the model fit statistics of this accepted model met all the requirements adopted in the current study to determine whether a model was accepted. The result indicated that this accepted measurement model of reading and test-taking strategy use generally depicted the data well. The accepted measurement model of reading and test-taking strategy use for the HEA group is presented in the following figure.



MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ= Monitoring and utilizing test questions; EM=Evaluating and marking; MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA= Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP=Marking incomprehensible parts; MKPO=Marking key points or options.

Figure 2 The final measurement model of reading and test-taking strategy use: The HEA group

The factor loadings ranged from .513 to .822 (see Figure 2), except the three cross-loadings. This suggested that observed variables (strategy subgroups) respectively well explained their latent variables (i.e., the MDAMT, the CME, the MUTQ, and the EM processes). Moreover, the figure also reveals that strategy use processes (i.e., the MDAMT, the CME, the MUTQ, and the EM processes) were correlated with each other, with correlation coefficients ranging from .310 to .635, not close to 1.000. This indicated that reading and test-taking strategy use was represented appropriately by these components (i.e., the MDAMT, the CME, the MUTQ, and the EM processes) based on the HEA group data.

Appendix 15

The Full Latent Variable Model regarding the Relationship among English Language Knowledge, Reading and Test-taking Strategy Use, and Multiple-choice Reading Comprehension Test Performance: The HEA Group

The accepted full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the HEA group is presented in the following figure.

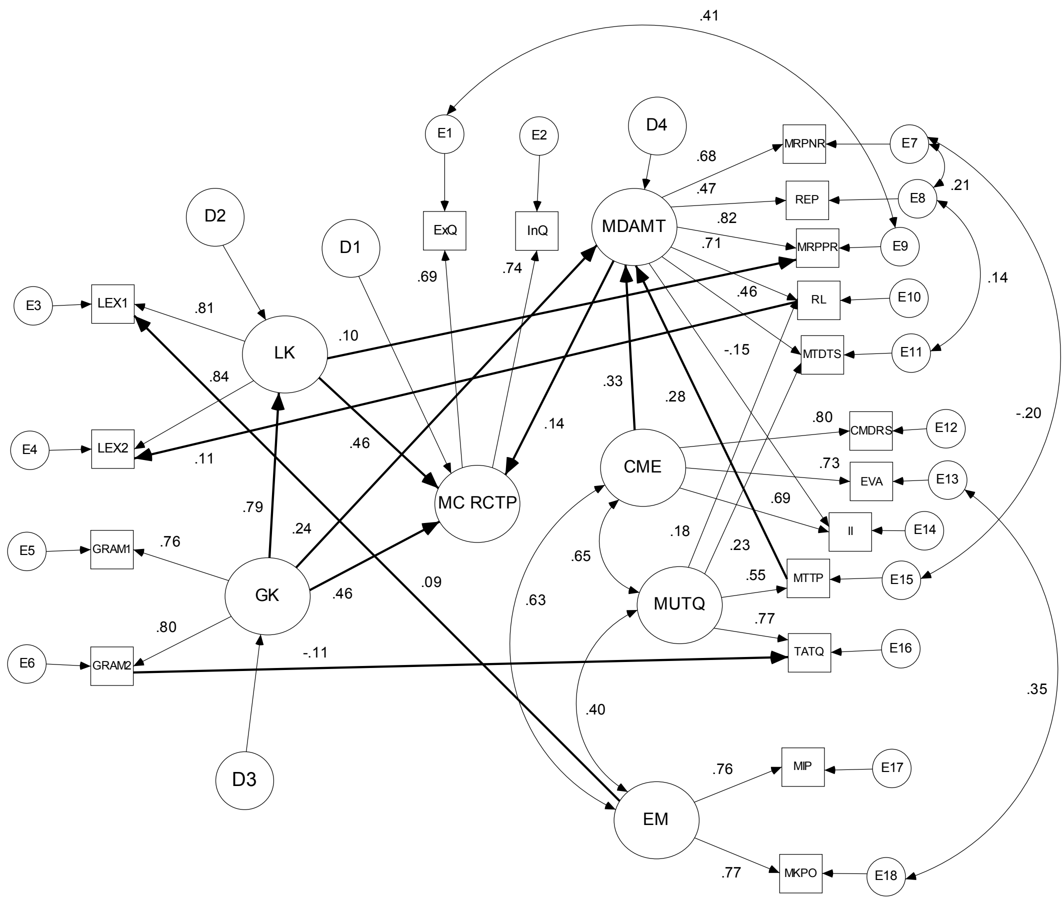


Figure 1 The final full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance: The HEA group

The covariance matrix for the full latent variable model regarding the relation among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the HEA group is offered as follows.

Table 4 The covariance matrix for the HEA group (Part I)

	MTTP	GRAM2	RL	InQ	ExQ	GRAM1
MTTP	4.681					
GRAM2	.000	3.817				
RL	1.529	.494	3.448			
InQ	.187	2.051	.714	4.256		
ExQ	.131	1.432	.498	1.639	2.401	
GRAM1	.000	2.039	.413	1.716	1.197	2.915
LEX1	.097	2.154	.536	2.301	1.606	1.802
LEX2	.187	2.162	.848	2.321	1.620	1.808
MKPO	.982	.000	1.011	.122	.085	.000
MIP	.887	.000	.913	.110	.077	.000
TATQ	4.805	-1.085	3.152	-.259	-.181	-.579
II	1.234	-.173	.993	-.041	-.029	-.144
EVA	1.328	.000	1.351	.162	.113	.000
CMDRS	3.027	.000	3.080	.370	.258	.000
MTDTS	1.730	.448	2.209	.670	.468	.374
MRPPR	1.106	.668	1.897	.868	.996	.559
REP	.717	.312	1.204	.434	.303	.261
MRPNR	.792	.643	2.364	.875	.611	.538

Table 4 The covariance matrix for the HEA group (Part II)

	LEX1	LEX2	MKPO	MIP	TATQ	II
LEX1	4.657					
LEX2	3.071	4.196				
MKPO	.416	.124	7.205			
MIP	.376	.112	3.822	6.003		
TATQ	-.286	-.211	3.316	2.995	27.479	
II	.093	-.027	2.495	2.254	4.872	9.545
EVA	.221	.165	3.220	2.025	4.485	3.303
CMDRS	.503	.377	5.110	4.615	10.222	7.527
MTDTS	.509	.656	1.154	1.042	4.020	1.246
MRPPR	.763	.907	.721	.651	1.723	.569
REP	.322	.416	.467	.422	1.150	.377
MRPNR	.662	.843	.961	.868	2.369	.818

Table 4 The covariance matrix for the HEA group (Part III)

	EVA	CMDRS	MTDTS	MRPPR	REP	MRPNR
EVA	5.633					
CMDRS	6.767	24.333				
MTDTS	1.546	3.524	6.799			
MRPPR	.959	2.185	1.856	2.488		
REP	.621	1.416	1.654	1.099	3.095	
MRPNR	1.279	2.916	2.326	2.150	1.975	6.304

Appendix 16
Confirmatory Factor Analysis for the Two-component Measurement Model of English Language Knowledge: The LEA Group

Based on the data collected from the LEA group, the proposed two-component measurement model of English language knowledge was tested by confirmatory factor analysis (CFA) with the use of structural equation modeling (SEM) procedures in which a *confirmatory modeling strategy* was adopted. Previous to performing SEM, I examined the z-score of each variable to identify the possible outlier (i.e., the case with the absolute value of the z-score greater than 3.000). Three cases were located and they were dropped. In addition, I inspected the multivariate normality to ensure that the data set would be generally multivariately distributed and then the parameter estimates estimated by the maximum-likelihood estimation procedures would not be impinged upon. The following table presents the result of the multivariate normality assessment.

Table 1 The assessment of the multivariate normality for the two-component measurement model of English language knowledge for CFA: The LEA group

Variable	min	max	skew	c.r.	kurtosis	c.r.
GRAM1	.000	8.000	-.207	-1.910	-.637	-2.934
GRAM2	.000	8.000	.177	1.628	-.747	-3.442
LEX1	1.000	10.000	-.385	-3.545	-.821	-3.779
LEX2	2.000	11.000	-.665	-6.127	-.454	-2.090
Multivariate					**-1.712**	-2.787

Note. n=509.

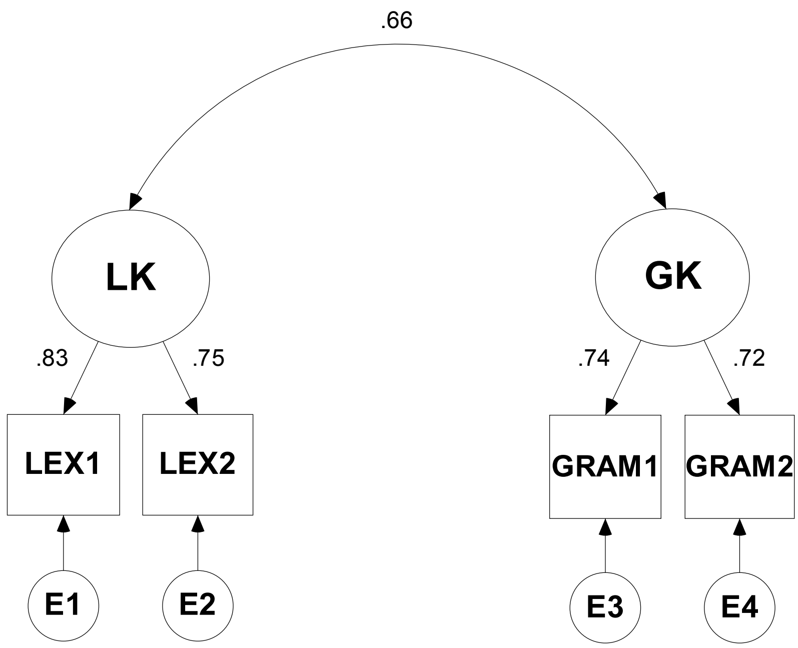
The absolute multivariate kurtosis value was 1.712 (see Table 1), which was within the acceptable limit (< 10). This suggested that the multivariate normality assumption was generally observed. Then, I carried out SEM to test the hypothesized measurement model of English language knowledge. Table 2 lists the summary of the evaluation of model fit.

Table 2 Summary of the evaluation of model fit for the two-component model of English language knowledge: The LEA group

Model fit indices	Levels of acceptable fit	Evaluation results
χ^2	Nonsignificant with the p-value > .050	Poor (χ^2 = 4.009, p = .045)
GFI	> .900	**Very good** (GFI = .996)
AGFI	> .900	**Very good** (AGFI = .961)
CFI	> .950	**Very good** (CFI = .995)
TLI	> .950	**Good** (TLI = .967)
RMSEA	< .060	Poor (RMSEA = .077)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

The above table shows that the model fit statistics of this accepted model satisfied all the requirements adopted in the current study to determine whether a model was accepted, except one statistic. This was excusable, given the value of the RMSEA approached the .060 cut-off. With the evaluation of the above model fit statistics, it can be concluded that this accepted measurement model of English language knowledge, generally, described the gathered data appropriately. The accepted measurement model of English language knowledge for the LEA group is presented in the following figure.



LK=Lexical knowledge; GK=Grammatical knowledge. LEX1 consists of ten test items of the vocabulary subtest; LEX2 eleven test items of the vocabulary subtest; GRAM1 eight test items of the grammar subtest; GRAM2 eight test items of the grammar subtest.

Figure 1 The measurement model of English language knowledge: The LEA group

The factor loadings ranged from .721 to .831 (see Figure 1). This suggested that observed variables (LEX1, LEX2, GRAM1 and GRAM2) respectively well explained their latent variables – lexical knowledge (LK) and grammatical knowledge (GK). Further, the figure displays that GK and LK were strongly correlated with each other, with a correlation coefficient of .660, not close to 1.000. This indicated that English language knowledge was represented appropriately by two components (i.e., lexical knowledge and grammatical knowledge) based on the LEA group data.

Appendix 17

Confirmatory Factor Analysis for the Four-component Measurement Model of Reading and Test-taking Strategy Use: The LEA group

Based on data gathered from the LEA group, the proposed four-component measurement model of reading and test-taking strategy use was tested by confirmatory factor analysis (CFA) with the use of structural equation modeling (SEM) procedures in which a *model generating strategy* was adopted. Prior to performing SEM, I inspected the z-score of each variable to identify the possible outlier (i.e., the case with the absolute value of the z-score greater than 3.000). Twenty-eight cases were located and they were deleted. In addition, I examined the multivariate normality to ensure that the data set would be generally multivariately distributed and then the parameter estimates estimated by the maximum-likelihood estimation procedures would not be impinged upon. The following table presents the result of the multivariate normality assessment.

Table 1 The assessment of the multivariate normality for the four-component measurement model of reading and test-taking strategy use: The LEA group

Variable	min	max	skew	c.r.	kurtosis	c.r.
MKPO	.000	15.000	-.200	-1.792	-.497	-2.231
MIP	.000	10.000	-.285	-2.561	-.818	-3.673
MTDTS	11.000	25.000	-.362	-3.248	.023	.105
RL	9.000	20.000	-.187	-1.683	-.228	-1.025
MRPPR	6.000	15.000	-.185	-1.663	-.142	-.639
REP	7.000	15.000	-.599	-5.381	-.317	-1.423
MRPNR	12.000	25.000	-.378	-3.395	-.180	-.810
TATQ	8.000	33.000	-.110	-.992	-.343	-1.540
MTTP	.000	10.000	-.393	-3.527	-.399	-1.790
II	.000	15.000	.278	2.501	-.453	-2.036
EVA	1.000	15.000	-.097	-.870	-.199	-.893
CMDRS	8.000	33.000	-.048	-.432	-.334	-1.499
Multivariate					**14.476**	8.687

Note. n=484.

The value for the multivariate kurtosis was 14.476 (see Table 1), which exceeded the acceptable limit (< 10). This suggested the apparent non-multivariate normality of the data. Based on the result of the Mahalanobis-d-squared, I dropped 13 cases of which the value of the Mahalanobis-d-squared exceeded 27. The sample ended up with 471. The following table presents the result of the multivariate normality assessment after 13 cases were dropped.

Table 2 The assessment of the multivariate normality for the four-component measurement model of reading and test-taking strategy use for CFA after 13 cases were dropped: The LEA group

Variable	min	max	skew	c.r.	kurtosis	c.r.
MKPO	.000	15.000	-.173	-1.535	-.515	-2.281
MIP	.000	10.000	-.290	-2.566	-.799	-3.539
MTDTS	11.000	25.000	-.375	-3.327	.062	.273
RL	9.000	20.000	-.145	-1.284	-.283	-1.252
MRPPR	6.000	15.000	-.171	-1.517	-.136	-.602
REP	7.000	15.000	-.578	-5.120	-.335	-1.482
MRPNR	12.000	25.000	-.347	-3.073	-.194	-.861
TATQ	8.000	33.000	-.081	-.721	-.377	-1.668
MTTP	.000	10.000	-.388	-3.435	-.397	-1.758
II	.000	15.000	.262	2.322	-.467	-2.071
EVA	1.000	15.000	-.087	-.770	-.215	-.954
CMDRS	8.000	33.000	-.029	-.253	-.349	-1.545
Multivariate					**9.160**	5.422

Note. n=471.

The absolute multivariate kurtosis value was 9.160 (see Table 2), which was within the acceptable limit (< 10). This indicated that the multivariate normality assumption generally was observed. Then, I conducted SEM to test the hypothesized measurement model of reading and test-taking strategy use. At first, the model was postulated as follows.

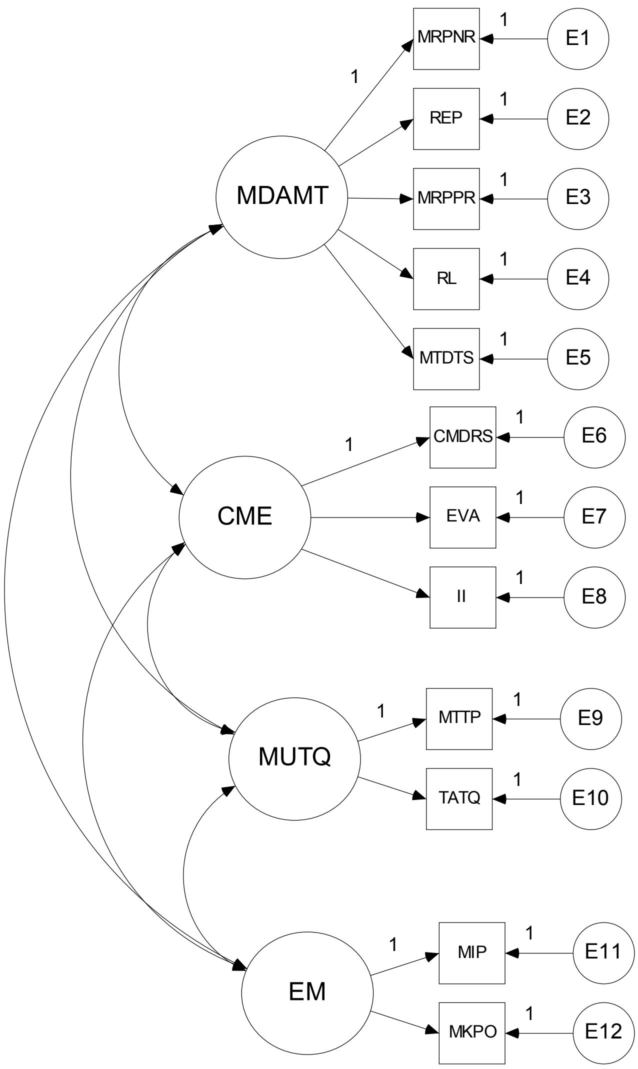


Figure 1 The original hypothesized measurement model of reading and test-taking strategy use: The LEA group

When examining the model fit statistics, I found that the statistics of the GFI (.956), the AGFI (.928) and the CFI (.959) were all above the cut-off value. The result implied that this measurement model appeared to describe the collected data well. However, I noticed that the value of the TLI (.944) were below the .950 cut-off, the chi-square statistic (χ^2= 127.304) was large, and the variance estimate of E12 was nonsignificant. Additionally, the factor loading that MKPO produced on EM was as high as 1.000, which was unreasonable in terms of the SEM analysis. Accordingly, I made an array of adjustments for and respecified the model, grounded on modification indices, interpretability, and the model of reading and test-taking strategy use for the entire group.

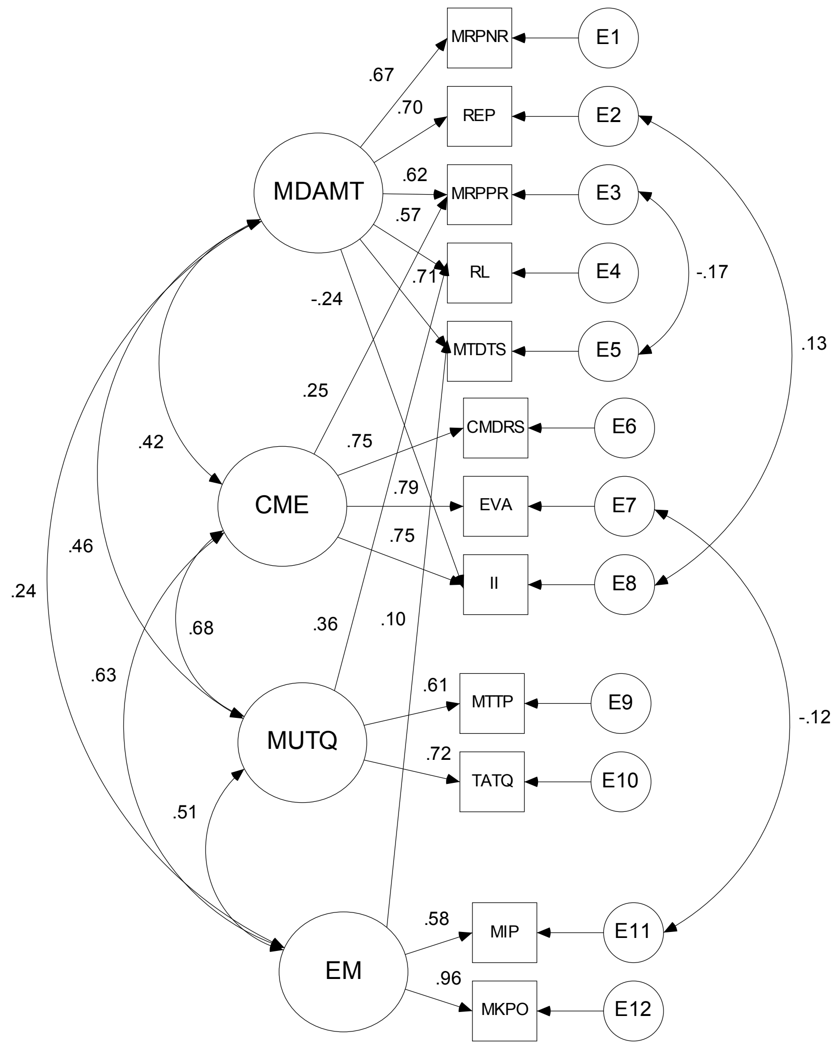
Finally, a model fitting the sample data well was produced. Table 3 provides the summary of the evaluation of model fit for this accepted model.

Table 3 Summary of the evaluation of model fit for the four-component model of reading and test-taking strategy use: The LEA group

Model fit indices	Levels of acceptable fit	Evaluation results
χ^2	Nonsignificant with the p-value > .050	**Good** (χ^2 = 37.484, p = .628)
GFI	> .900	**Very good** (GFI = .987)
AGFI	> .900	**Very good** (AGFI = .975)
CFI	> .950	**Very good** (CFI = 1.000)
TLI	> .950	**Very good** (TLI = 1.003)
RMSEA	< .060	**Very good** (RMSEA = .000)

Note. GFI=The goodness-of-fit index; AGFI=The adjusted goodness-of-fit index; CFI=The comparative fit index; TLI=The Tucker-Lewis index; RMSEA=The root mean square error of approximation.

The above table shows the chi-square (χ^2) statistic of 37.484 – even smaller than the previous one. The model fit statistics of this accepted model satisfied all the requirements adopted in the current study to determine whether a model was accepted. This indicated that this accepted measurement model of reading and test-taking strategy use depicted the data appropriately. The accepted measurement model of reading and test-taking strategy use for the LEA group is presented in the following figure.



MDAMT=Monitoring, directing attention and managing the test; CME=Constructing the meaning and evaluating; MUTQ= Monitoring and utilizing test questions; EM=Evaluating and marking; MRPNR=Monitoring the reading process with negative results; REP=Repeating; MRPPR=Monitoring the reading process with positive results; RL=Retrieving-linking; MTDTS=Managing the test with the deployment of test-taking strategies. CMDRS=Constructing the meaning with the deployment of reading strategies; EVA= Evaluating; II=Interacting with the input. MTTP=Monitoring the test-taking process; TATQ=Taking advantage of test questions. MIP=Marking incomprehensible parts; MKPO=Marking key points or options.

Figure 2 The final measurement model of reading and test-taking strategy use: The LEA group

The factor loadings ranged from .569 to .956 (see Figure 2), except the four cross-loadings. This suggested that observed variables (strategy subgroups) respectively well explained their latent variables (i.e., the MDAMT, the CME, the MUTQ, and the EM processes). Furthermore, the figure displays that strategy use processes (i.e., the MDAMT, the CME, the MUTQ, and the EM processes) were correlated with each other, with correlation coefficients ranging from .238 to .683, not close to 1.000. This indicated that reading and test-taking strategy use was represented appropriately by these components (i.e., the MDAMT, the CME, the MUTQ, and the EM processes) based on the LEA group data.

Appendix 18

The Full Latent Variable Model regarding the Relationship among English Language Knowledge, Reading and Test-taking Strategy Use, and Multiple-choice Reading Comprehension Test Performance: The LEA Group

The accepted full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the LEA group is presented in the following figure.

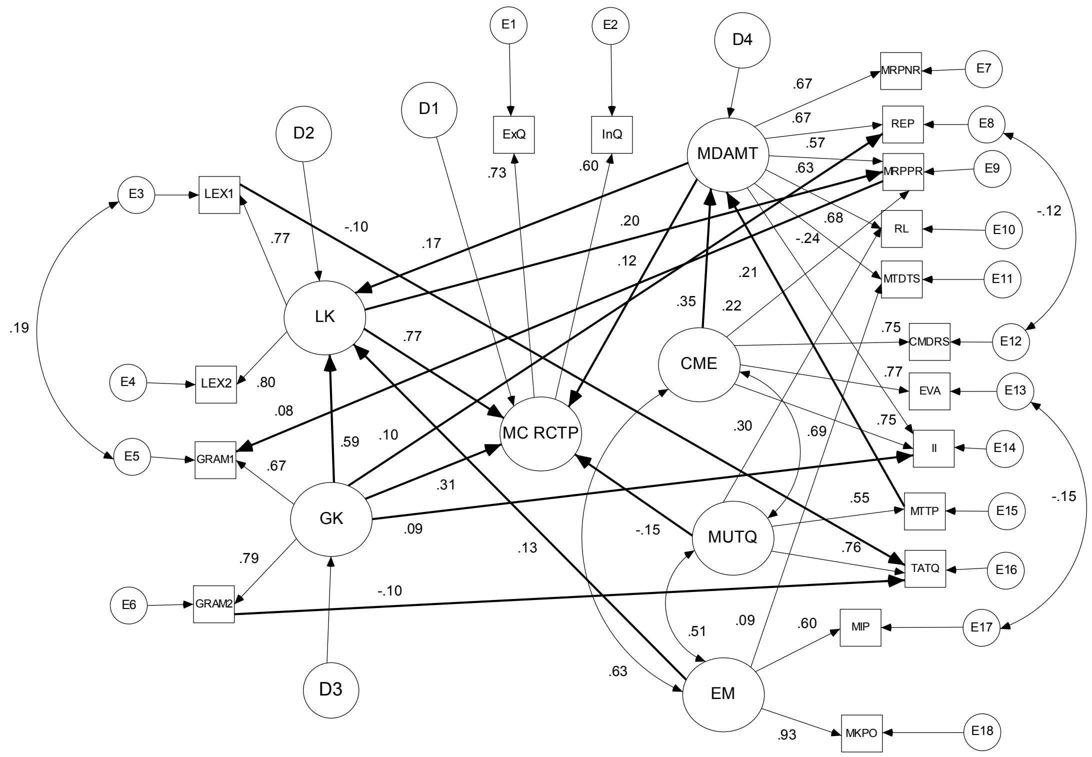


Figure 1 The final full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance: The LEA group

The covariance matrix for the full latent variable model regarding the relation among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the LEA group is provided as follows.

Table 4 The covariance matrix for the LEA group (Part I)

	MTTP	GRAM2	LEX1	MRPPR	InQ	ExQ
MTTP	5.400					
GRAM2	.000	4.005				
LEX1	.386	1.640	5.225			
MRPPR	1.300	.343	1.139	3.530		
InQ	.083	1.377	1.913	.701	3.539	
ExQ	.098	1.627	2.260	.828	1.514	3.326
GRAM1	.105	1.961	1.733	.553	1.127	1.332
LEX2	.399	1.696	3.186	1.178	1.978	2.337
MKPO	1.793	.000	.840	1.710	.285	.336
MIP	1.042	.000	.488	.994	.165	.195
TATQ	4.830	-1.332	-.688	2.382	-.796	-.941
II	1.539	.468	.690	1.322	.248	.293
EVA	1.747	.000	.535	1.839	.150	.178
CMDRS	3.222	.000	.986	3.391	.277	.327
MTDTS	1.736	.000	.780	2.721	.521	.615
RL	2.031	.000	.667	2.366	.355	.419
REP	1.087	.327	.702	1.816	.523	.617
MRPNR	1.470	.000	.657	2.395	.462	.546

Table 4 The covariance matrix for the LEA group (Part II)

	GRAM1	LEX2	MKPO	MIP	TATQ	II
GRAM1	3.338					
LEX2	1.414	5.149				
MKPO	.139	.869	8.642			
MIP	.080	.505	4.380	7.160		
TATQ	-.633	-.232	5.159	2.998	25.292	
II	.471	.714	3.666	2.130	5.121	10.687
EVA	.149	.553	3.442	1.467	5.089	4.220
CMDRS	.275	1.019	6.348	3.689	9.385	7.782
MTDTS	.220	.807	2.184	1.269	2.990	.887
RL	.192	.690	2.059	1.197	4.430	1.370
REP	.401	.726	1.042	.605	1.617	.448
MRPNR	.194	.680	1.409	.819	2.358	.522

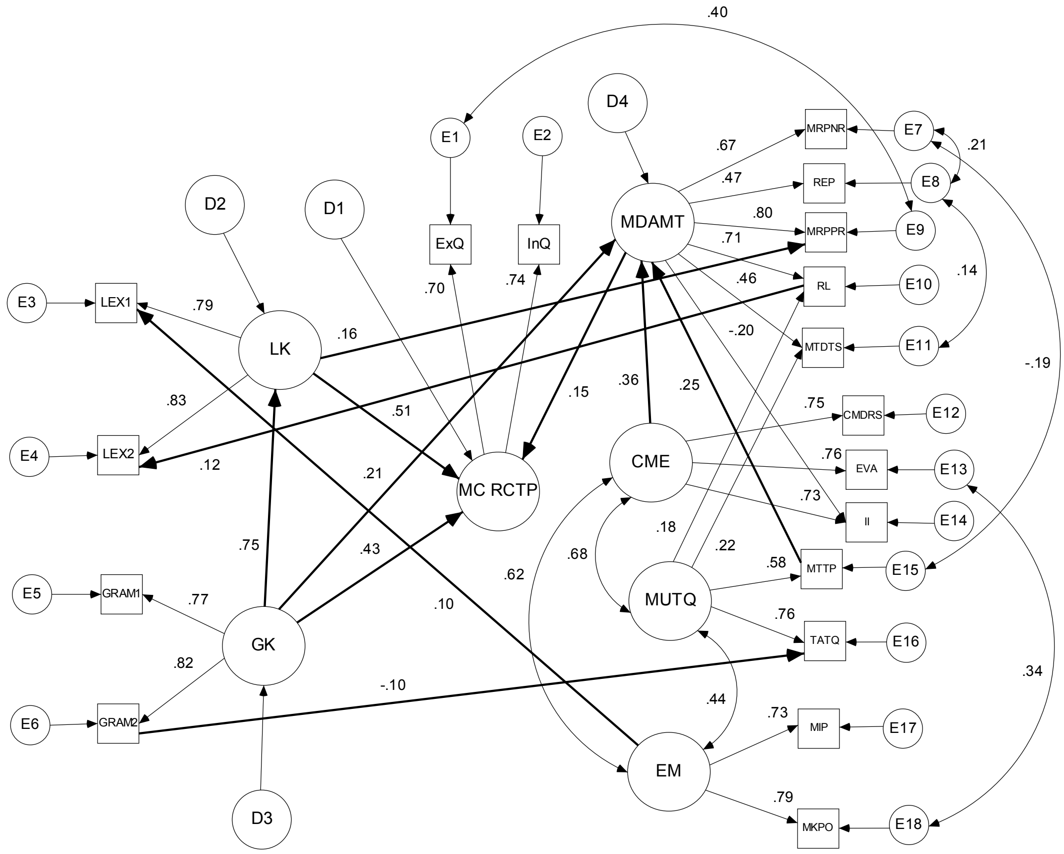
Table 4 The covariance matrix for the LEA group (Part III)

	EVA	CMDRS	MTDTS	RL	REP	MRPNR
EVA	6.616					
CMDRS	7.269	24.104				
MTDTS	1.982	3.656	8.199			
RL	2.156	3.975	3.454	5.115		
REP	1.155	1.577	2.719	2.254	4.000	
MRPNR	1.562	2.881	3.677	3.049	2.449	7.334

Appendix 19

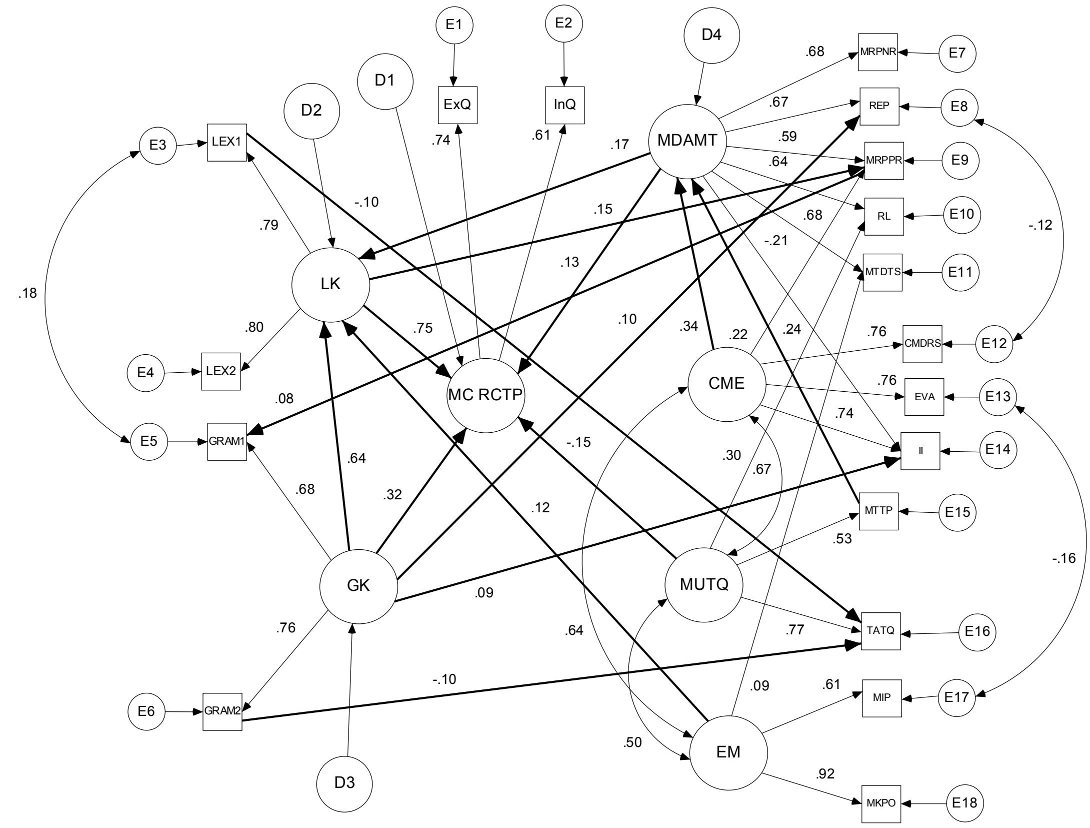
The Full Latent Variable Models regarding the Relationship among English Language Knowledge, Reading and Test-taking Strategy Use, and Multiple-choice Reading Comprehension Test Performance for the HEA Group and the LEA Group: A Simultaneous Group Analysis

The full latent variable model accepted in the simultaneous group analysis is presented in the following figure.



Between the HEA group and the LEA group, equality constraints were imposed on LK→LEX1, GK→GRAM1, MDAMT→RL, MDAMT→II, CME→EVA, CME→II, MUTQ→TATQ, MDAMT→MC RCTP, CME→MDAMT, CHW→MDAMT, GK→MC RCTP, GK→LK, GRAM2→TATQ, LK→MRPPR, CME←→MUTQ, CME←→EM, MUTQ←→EM.

Figure 1 The full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the HEA group: The simultaneous group analysis



Between the HEA group and the LEA group, equality constraints were imposed on LK→LEX1, GK→GRAM1, MDAMT→RL, MDAMT→II, CME→EVA, CME→II, MUTQ→TATQ, MDAMT→MC RCTP, CME→MDAMT, CHW→MDAMT, GK→MC RCTP, GK→LK, GRAM2→TATQ, LK→MRPPR, CME←→MUTQ, CME←→EM, MUTQ←→EM.

Figure 2 The full latent variable model regarding the relationship among English language knowledge, reading and test-taking strategy use, and multiple-choice reading comprehension test performance for the LEA group: The simultaneous group analysis