

國立台灣師範大學英語學系

碩 士 論 文

Master's Thesis

Graduate Institute of English

National Taiwan Normal University

英語閱讀測驗之內涵與閱讀能力之評析：以大學
學科能力測驗與指定科目考試為例

An Analysis of the Reading Skills Measured in
Reading Comprehension Tests on the Scholastic
Achievement English Test (SAET) and the
Department Required English Test (DRET)

指導教授：陳秋蘭博士

Advisor: Dr. Chiou-lan Chern

研究生：陳秀娟撰

Hsiu-chuan Chen

中華民國九十八年七月

July 2009

摘要

本研究使用語言學家 Nuttall 的閱讀技巧分類之修正版來探討近六年（2002-2007 年）大學學科能力測驗及指定科目考試英文閱讀測驗試題中欲測試的閱讀技巧為何，以及歷屆考生在各類題型上的表現。

本研究採質化與量化分析法。質的分析採內容分析法，將一百六十七個考題依照 Nuttall 的閱讀技巧分類之修正版分類。量化分析則使用電腦統計軟體 SPSS 13.0 來檢測各種閱讀技巧類型在考試中出現的頻率及分布，雙因子變異數分析用來個別檢測二種考試中各類題型之答對率間有無顯著差距及其差距是否每年相同，同時並了解高分組與低分組學生於各類題型上表現的差距。

本研究之主要研究發現如下：

（一）研究結果顯示，在二種考試中，試題欲測量的閱讀技巧可以分為八類（「由上下文推測字彙意義」、「辨認連接詞」、「辨認細節」、「辨認功能價值」、「辨認文章組織結構」、「辨認作者的預設立場」、「推論」以及「辨認主旨」。

（二）「辨認細節」這類考「由下往上」(bottom-up)的技巧的題目類型在二種考試中最常被考，因此可推論二種考試都比較偏好考「由下往上」(bottom-up)類型的技巧，而出現次數最少的題型則為「辨認文章組織結構」。此外，大學學科能力測驗及指定科目考試的英文閱讀測驗的最大差異在於閱讀技巧類型出現的頻率、出現處以及分佈。在大學學科能力測驗中每年都會出現的技巧是「由上下文推測字彙意義」以及「辨認細節」此二種題型，而在指定科目考試的英文閱讀測驗每年都出現的則只有「辨認細節」此類題型。

（三）二因子變異的分析研究顯示在大學學科能力測驗以及指定科目考試中，不同題型對於答對率並沒有顯著影響，亦即不同題型答對率的高低在六年當中並不一致。

（四）在二種考試當中，低層次的閱讀技巧(local skill)最能區隔高低分組的學生。在大學學科能力測驗當中，所有的題型在高低分組的答對率差距方面都有達到最低標準，然而在指定科目考試方面有二類題型的鑑別度低於最低標準，像

是「推論」以及「辨認作者的預設立場」。此結果暗示這二類題型可能對所有的考生來說太難，以致於無法適當區隔高低分組的表現。

根據上述分析結果，本研究最後提出一些教學建議以供參考。

Abstract

The present study aimed to adopt a revised version of Nuttall's taxonomy to investigate the reading skills measured in the SAET (Scholastic Achievement English Test) and the DRET (Department Required English Test) administered from 2002 to 2007, and to explore how test takers (all examinees, high achievers, and low achievers) performed on different types of items.

Both qualitative and quantitative analyses were adopted. The qualitative analysis was conducted by categorizing each of the 167 reading comprehension items into reading skill type in the revised Nuttall's Taxonomy. For the quantitative analysis, SPSS 13.0 statistical package was used to examine the frequency distribution of the item types. The two-way ANOVA test was applied to the SAET and the DRET to see whether there were significant differences among the passing rates of different question types and to investigate whether these differences were consistent throughout the years. Also, the discrimination indexes were analyzed via the two-way ANOVA to see how the high achievers and low achievers differed while answering different types of reading questions each year. The results of this study are summarized as follows:

First, the findings showed that in both tests, eight types of reading skills were measured: "Word Inference from Context," "Recognizing Cohesive Devices," "Recognizing and Interpreting Details," "Recognizing Functional Value," "Recognizing Text Organization," "Recognizing Presuppositions Underlying the Text," "Recognizing Implications and Making Inferences," and "Recognizing and Understanding the Main Idea."

Second, the most frequent items tested in the SAET and DRET are items on "Recognizing and Interpreting Details," which indicated that this type of reading skill is favored in both tests. However, "Recognizing Text Organization" is the least tested skill. In addition, the similarities and differences between the SAET and DRET lay in the frequency, occurrences, and distribution of reading skill item types. Two types of items occurred every year in the SAET, including local items on "Word Inference from Context" and "Recognizing and Interpreting Details." However, only "Recognizing and Interpreting Details" occurred every year in the DRET.

Third, the ANOVA analysis showed that there was no significant effect of items

on the examinees' average passing rates. In other words, the ranking of passing rates of different item types in the SAET and DRET were not consistent throughout the years.

Finally, in both SAET and DRET, items on local skills best discriminated high and low achievers. In the SAET, the discrimination indexes of all item types reached the ideal discrimination index whereas in the DRET two types of items had unsatisfactory discriminatory power: items on "Recognizing Implications and Making Inferences" and "Recognizing Presuppositions Underlying the Text." This indicated that these two types of items were probably too difficult for most examinees and did not appropriately distinguish the high and low achievers.

Based on the aforementioned analysis of results, some pedagogical implications for reading instruction and testing in senior high schools were provided.

Acknowledgements

I owe my deepest gratitude to my advisor Dr. Chou-lan Chern, chair of Department of English of National Taiwan Normal University, whose inspiring guidance, encouragement, patience and support helped me in all the time of research and writing of this thesis. I would also like to extend my heart-felt thanks to Dr. Hsi-chin Chu and Dr. Chin-kuei Cheng for serving on my thesis committee and for their valuable comments, suggestions, and advice.

I would like to thank all the faculty members in the English Department for making my Master's program a good experience. Also, I am very grateful to my family back home for always standing by my side. I am thankful to all my friends and my boyfriend Tsung and his family who always have supported and encouraged me. Finally, I would like to extend my warmest thanks to two wonderful teachers, Elaine and Julie, for being my raters and for their support and encouragement.

TABLE OF CONTENTS

	Page
CHINESE ABSTRACT.....	i
ENGLISH ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	vii
CHAPTER ONE INTRODUCTION.....	1
Background and Motivation.....	1
Research Questions	5
Significance of the Present Study.....	6
Organization of the Thesis.....	6
CHAPTER TWO LITERATURE REVIEW	9
The Rationale of Reading	9
Reading Process Models	9
Reading Comprehension and Reading Skills.....	14
Theoretical Reading Skill Frameworks.....	15
The Original Nuttall’s Taxonomy.....	17
Reading Skill Instruction.....	24
Testing ESL/EFL Reading Comprehension.....	26
Why Testing (The Relationship of Teaching and Testing).....	27
Testing Comprehension with Multiple-choice Questions.....	28
Studies of Reading Tests Analysis (Item Analysis).....	30

Summary	38
CHAPTER THREE METHOD.....	41
Materials.....	41
Participants.....	42
Instrument.....	43
Data Analysis.....	46
Formal Analysis.....	46
Analysis of the Data Coded and the Passing Rates.....	48
Summary.....	49
CHAPTER FOUR DATA ANALYSIS AND RESULTS.....	51
Qualitative and Quantitative Data Analysis and Results.....	51
The Reading Skills Measured in SAET & DRET.....	51
Similarities and Differences between the SAET & the DRET.....	53
Similarities and Differences between Reading Skills Measured in the SAET and the DRET.....	55
Examinees' Performance on Each Question Type	59
SAET.....	59
DRET.....	63
Comparisons of High and Low Achievers' Performances on Different Item Types.....	66
Results of the SAET.....	67
Results of the SAET.....	74
Summary	80

CHAPTER FIVE DISCUSSION AND CONCLUSION.....	83
Discussion of Major Findings.....	83
Reading Skills Measured in the SAET & DRET.....	84
Similarities and Differences between the Reading Skills Tested in the SAET and the DRET.....	84
Examinees' Performances on Various Item Types.....	86
Comparisons of High and Low Achievers' Performances on Different Item Types.....	88
Conclusions.....	90
Summary of the Major findings.....	90
Pedagogical Implications.....	91
Limitation Suggestions for Future Research	93
REFERENCES.....	95
APPENDIX 1: Principles for Classification of the Reading Skills	
Based on the Revised Nuttall's Taxonomy.....	102
APPENDIX 2: Results of the Formal Item Analysis.....	107

LIST OF TABLES

Table 1 Structure of Nuttall' Original Taxonomy	22
Table 2 The Number of Reading Passages and Comprehension Test Items	42
Table 3 The Number of Examinees Taking the SAET from 2002 to 2007	42
Table 4 The Number of Examinees Taking the DRET from 2002 to 2007	43
Table 5 Structure of the Revised Nuttall's Taxonomy	45
Table 6 Reading Level Tested in the SAET & DRET	51
Table 7 The Frequencies of Reading Skills Measured in the SAET & the DRET	52
Table 8 Reading Level Tested in the SAET & DRET from 2002 to 2007	53
Table 9 Reading Skills Measured in both SAET & DR ET	54
Table 10 Reading Skills Measured in 2002-2007 SAET & DRET	58
Table 11 Average Passing Rates of Items Measuring Different Reading Skills in 2002-2007 SAET and DRET	60
Table 12 The ANOVA Analysis of the Passing Rates in the SAET	63
Table 13 The ANOVA Analysis of the Passing Rates in the DRET	66
Table 14 Passing Rates of High & Low Achievers and the Discrimination Indexes on Different Reading Levels in the SAET & DRET	68
Table 15 Passing Rates of the High & Low Achievers and the Discrimination Indexes on Different Item Types in 2002-2007 SAET	69
Table 16 The ANOVA Analysis of Discrimination Indexes in the SAET	74
Table 17 Passing Rates of the High & Low Achievers and the Discrimination Indexes on Different Item Types in 2002-2007 DRET	76
Table 18 The ANOVA Analysis of Discrimination Indexes in the DRET	79



CHAPTER ONE

INTRODUCTION

Background and Motivation

English reading is an essential skill for those who use English as a second or foreign language. It is probably by far the most important of the four skills (i.e., listening, speaking, reading and writing) for English as a second language (ESL) & English as a foreign language (EFL) learners (Carrell, 1988a). For senior high school students in the EFL context of Taiwan, English reading ability is of paramount importance since English is taught as an academic subject in the school setting. In addition, a good command of English reading is important in passing nationwide English tests such as the Scholastic Achievement Test (SAT)¹ and the Department Required Test (DRT)². The College Entrance Examination Center (hence CEEC) implements a two-stage nationwide examination every year to determine who goes to which university. Therefore, both tests have a huge impact on teaching and learning for senior high school teachers and students in Taiwan.

The SAT is an achievement test aiming to assess senior high school students' basic scholastic knowledge and abilities. It is designed to choose potential candidates for universities at an early stage. The DRT, on the other hand, serves as a proficiency test to identify competent candidates who possess certain abilities to do well in certain subject areas required by particular universities. Usually, students need the results of tests on three to six different subjects to apply for schools of their choice. The goals of

¹ The SAT was firstly administered in 1994.

² The DRT was originally called the Joint College Entrance Examination (JCEE) before 2002. It was first implemented in 2002 when the new college admission policy was introduced.

the SAT are: (1) to see whether the candidates have the abilities required of senior high school students in fundamental subjects, (2) to see whether candidates have acquired abilities for college education, (3) to see whether candidates have abilities to combine knowledge they learn at school with real-life experience or to integrate it with knowledge in other fields, and (4) to see whether candidates have the abilities to comprehend and apply knowledge. On the other hand the goals of the DRT are (1) to see whether candidates understand the important knowledge in the discipline, (2) to see whether candidates have the abilities to read, to judge, to infer, and to analyze materials, (3) to see whether candidates have abilities to express their ideas, and (4) to see whether candidates have abilities to apply the knowledge in the disciplines. Comparing the purposes and goals of the SAT and DRT, the major difference between these two tests is that the SAT focuses more on measuring students' basic abilities whereas the DRT emphasizes more on measuring the students' higher-order cognitive abilities—to judge, to infer and to analyze materials. Thus, the DRET (i.e., Department Required English Test), with its aim to discriminate more proficient students in the English subject, is relatively more difficult than the SAET (i.e., Scholastic Achievement English Test).

Both Scholastic Achievement English Test and Department Required English Test play vital roles in university admission and both tests exert considerable influence on English teaching and learning. As a result, what is measured in the reading comprehension section in both tests has drawn interest among teachers and students (Lu, 2002). On both tests, the reading comprehension section, which usually occupies the largest proportion of scores in the whole English test, is a major component used to assess reading ability and the reading comprehension test items are assumed to tap different reading skills. Since both tests exert a great influence on

teaching and testing, both tests are analyzed annually to help teachers and students better understand the tests. Researchers have been interested in what should be and is tested in reading comprehension tests and have noted the importance of assessing reading skills. According to Cohen (1994), in addition to the types of meaning assessed by reading comprehension questions, test writers and users should be aware of the skills being tested. Alderson (in *Nuttall*, 1996, p. 219) points out that it is important to define the reading skills to be tested and to write questions which involve such skills. As Yu (2006) pointed out, senior high school teachers in Taiwan tend to see vocabulary teaching and structure analysis as the focus of reading instruction. In order to turn this around, the best way is probably to raise teachers' awareness of the concept and application of reading skills tested in the SAET or the DRET.

The CEEC makes an annual report on the item analysis of both SAET and DRET. The statistical analyses by the CEEC showed that most of the items had good statistical values, but test items with good statistical values might not necessary be good items (Huang, 1994). Also, the analysis has been focused mainly on the statistical values of items and yet in-depth qualitative analysis of the items which measure reading comprehension skills in SAET and DRET is rare. Most of the previous studies on the SAET and DRET aimed to provide an overview of the test construction or statistical results of item analysis rather than a thorough qualitative analysis on the reading skills tested on each comprehension test item in particular (e.g. CEEC Web site; Huang, 1994; Jiang & Lin, 1999; Jeng, 1992; Xu & Lu, 1998). In recent years, three researchers in Taiwan have more thorough research on the reading comprehension questions—one analyzed reading comprehension test items in the SAET (Lu, 2002), the other focused on the reading comprehension test items in the DRET (Hsu, 2004), and still another analyzed reading comprehension items in both SAET and DRET (Lan, 2007). However, the first two studies have been limited to the

use of the same taxonomy which is Mo's taxonomy (Mo, 1987). Mo's scheme (Mo, 1987), derived by summarizing six types of skills from both Carver's (1878) and Raygor & Raygor's (1985) taxonomies, appeared to be very general and failed to provide clear and detailed descriptions and examples of each skill. What's more, these two studies exhibited some problems with the categorization of test items. It was found that some of the categorizations of test items in both studies were inappropriate because Mo's taxonomy is too general, making the statistical analyses and the studies questionable. The third study by Lan (2007), adopted a revised Bloom's taxonomy to analyze the cognitive levels and knowledge types measured in both SAET and DRET. The taxonomy has proved to be useful in examining the cognitive levels and knowledge types of reading comprehension items; however, it is not specifically designed for ESL/EFL context. Both SAET and DRET are extremely important, yet previous studies concerning reading skills measured in SAET and DRET were far from conclusive. Only two taxonomies were used to examine the reading skills measured in these two tests and problems have been found with the two studies which analyzed the SAET and DRET by adopting Mo's taxonomy.

Nuttall's taxonomy is similar to Mo's taxonomy, but it does not have the problem exhibited in Mo's taxonomy. In Nuttall's taxonomy, the ability to recognize cohesive devices and the ability to recognize text organization are treated as separate skills. Therefore, Nuttall's taxonomy would be a better taxonomy in identifying the skill to recognize cohesive devices and the skill to recognize text organization. Moreover, Nuttall divided reading skills into bottom-up (lower-order) and top-down (higher order) skills. By adopting Nuttall's taxonomy, the researcher is able to find out whether global or local reading skills are more tested in the SAET and the DRET. Also, Nuttall's taxonomy contains skills that were measured either on the SAET or the DRET but items which measure these skills could not be identified by Mo's

taxonomy. For example, one skill on Nuttall's taxonomy—*Recognizing Functional Value*—had been measured on the DRET (See Appendix 1 for the principle and example question of this skill); however, by using Mo's taxonomy, items which test this skill were classified as items on inference. Thus, we can assume that Nuttall's taxonomy could reflect the skills tested but were not identified according to Mo's taxonomy. Nevertheless, a few limitations were found with Nuttall's taxonomy and therefore a revised version of Nuttall's taxonomy was formed for the present study. The original Nuttall's taxonomy is presented in Chapter Two and the revised taxonomy is presented in Chapter Three.

Though it is still undecided whether the reading ability can be measured by reading comprehension tests, so far no definite answer can be given to prove otherwise. Thus, the items in the reading tests are still worthy of analysis. The present study, viewing reading as a cognitive activity which can be broken down into component skills, attempts to apply a revised Nuttall's taxonomy to analyze the reading comprehension test items in both SAET and DRET from 2002 to 2007. The taxonomy adopted for the present study is specified in the chapter which describes the method.

Research Questions

The present study attempts to answer the following questions:

1. What are the reading skills measured in the SAET and DRET in the reading comprehension section from 2002~2007?
2. What are the similarities and differences between the SAET and DRET with regard to the reading skills measured?
3. How did high school students perform on test items that measure different skills?
4. How did high achievers and low achievers differ on test items that measure

different skills?

Significance of the Present Study

Over the past decades, several studies have been conducted with regard to the reading comprehension item analysis in SAET and DRET. Most of them focused on the statistical values of distractors, topics of passages, and so on. In view of the importance of the role reading skills play in reading instruction and assessment, the purpose of the present study is to examine the reading skills measured in the SAET and DRET so as to provide teachers and students with a better understanding of these two tests and to provide implications for reading instruction in senior high schools. This study, different from previous studies, attempts to investigate the reading skills measured in the reading comprehension questions in the SAET and DRET from 2002 to 2007 by adopting a taxonomy based on Nuttall's rubrics. It seeks to examine the reading skills measured in both tests, to see what reading skills were tested and to see if there's any difference between the two tests regarding the type of reading skills measured. The study is significant in that it provides both teachers and students in Taiwan with a better understanding of reading skills measured in SAET and DRET. It is hoped that the findings in the present study will help teachers gain some insights of the reading skills measured in both tests and help students develop reading skills appropriate for the exams.

Organization of the Thesis

The remainder of this study is structured as follows. Chapter Two deals with the literature review regarding the rationale of reading (e.g. reading process models), existing frameworks of reading skills, the original Nuttall's taxonomy, studies on reading skill instruction, and studies of reading comprehension tests. Chapter Three

describes the method and procedures for the collection of data on reading skill analyses. The fourth chapter presented the results, with a thorough description of the data analyses. Finally, results are discussed, conclusions are drawn with pedagogical implications, followed by limitations of the current study and suggestions for future study.

CHAPTER TWO

LITERATURE REVIEW

This study aimed to investigate the reading skills measured in the reading comprehension section in SAET and DRET. Thus, the theoretical foundations for the investigations of reading skills in tests are reviewed. First of all, section one explores the rationale of reading—the reading process models. Then, section two accounts for the relationship between reading comprehension and reading skills and existing reading taxonomies. Next, the third section presents Nuttall’s original taxonomy. Furthermore, the fourth section reviews studies on reading skill instruction. Finally, the last section discusses issues of assessing reading in an ESL/EFL context, including the advantages and disadvantages of the multiple-choice question technique and a review of recent studies on SAET and DRET.

The Rationale of Reading

Considerable research has been done to gain insights into the nature of reading and various theories exist to explain what is involved when we read. Generally speaking, reading is a cognitive activity since “it largely takes place in the mind” (Urquhart & Weir, 1998). To comprehend and interpret a text appropriately, a reader must be actively engaged in a mental process with the text. In order to look for ways to describe the interaction between the reader and the text, researchers have created models that describe what happens when people read. It is commonplace to classify reading process models into three types: the bottom-up model, the top-down model, and the interactive model. A general review of literature of these three models is provided as follows.

Reading Process Models

The Bottom-up Model. The bottom-up model, essentially data-driven, was

initially proposed by Gough (1972). This model involves the reader paying close attention to every detail of the text. In bottom-up processing, the meaning is arrived at, from bottom to top, where the reader begins with the printed text by decoding letters, words, phrases, and sentences and then builds up comprehension in a somewhat linear way from the incoming language data (Omaggio, 2001). The bottom-up model typically consists of lower-level processes in which the reader focuses on the processing of smaller linguistic units (e.g., letter, words) and then works towards larger units such as sentences, arriving at meaning from parts to whole. Therefore, in this view, readers are seen as passive identifiers of sequential graphic, phonemic, syntactic, and semantic systems of language (Alderson, 2000). According to Eskey (1973) (cited in *Carrell*, 1988a, p.3), this model was inadequate because it failed to recognize the readers' contribution during the reading process. It failed to recognize that the readers brought prior knowledge and expectations to bear on the task of understanding the written language.

The Top-down Model. In place of the bottom-up approach, Goodman (1967) offered a different kind of model which has been characterized as a top-down model of the reading process. Goodman (1967) claimed that reading is a “psycholinguistic guessing game” in which readers sample from the text³ and then confirm and correct their hypotheses about the predicted meaning of the text based on the message conveyed in the text. In other words, readers use guessing as a basis for comprehension and reconstruct meaning by predicting and confirming. By top-down processing, readers create meaning by relating to what the text says to what they already know. Smith (1971) and Goodman (1967) are major proponents of this model.

³ *Sampling* implies that the reader does not read every feature of every word, but rather focuses on certain features, certain words or certain phrases to get a sense of the meaning from the text.

In this view, readers become active participants in the reading process in which the readers make predictions and process information by utilizing their prior experience or background knowledge (Carrell, 1988b). Elaborating on Goodman's psycholinguistic model, Coady (1979) argued that reading comprehension required three components: background knowledge, higher-level conceptual abilities (e.g., intellectual capacity), and process strategies (e.g., grapheme-morpho-phoneme correspondences, syllable-morpheme information, syntactic information, lexical meaning, and contextual meaning).

The top-down (reader-driven or conceptually-driven) model, which involves the reader sampling text information to verify hypothesis and predictions, emphasizes the importance of the readers' schemata (Alderson, 2000). Carrell & Eisterhold (1983) claimed that there are two types of schemata—formal schemata, i.e., knowledge of language and linguistic conventions (e.g., knowledge of rhetorical structures, how a text is organized), and content schemata, which refers to knowledge of the world or knowledge about the content area of a text. The schemata the reader brings to the reading process is said to guide the reader's processing of text. According to schema theory, comprehending a text is an interactive process between the reader's background knowledge and the text, requiring the reader to activate relevant schemata to process and interpret new information quickly and efficiently (Richards & Schmidt, 2002).

The Interactive Model. By bottom-up processing only, readers rely too much on the lower-level skills, and by top-down processing only, readers rely too much on their prior knowledge. Thus, neither of these two models is an adequate characterization of the reading process (Alderson, 2001). More recent research on teaching reading has shown that the interactive model, a combination of top-down and bottom-up processing, is nearly always a prerequisite for successful reading

instruction (Brown, 2001). “Good reading — that is, fluent and accurate reading — can result only from a constant interaction between these processes (bottom-up & top-down)” (Eskey, 1988, p.95). The interactive model is a process model involving the combination and integration of bottom-up processing and top-down processing skills alternatively or at the same time (Carrell, 1988; Grabe, 1991; Nassaji, 2002). According to Rumelhart (1977, 2004), the interactive process model stresses that reading comprehension is reached when various sources of knowledge interact with sensory input, including orthographic knowledge, lexical knowledge, syntactic knowledge, semantic knowledge and pragmatic knowledge. Stanovich (1980), on the other hand, had proposed an interactive compensatory model. Stanovich’s model is interactive in that readers employ both bottom-up and top-down processing skills whereas it is compensatory in the sense that readers process texts by utilizing information from various resources of knowledge and a deficit in one knowledge source will result in a heavier reliance on other knowledge sources. In other words, readers with poor linguistic knowledge may use their top-down skills (e.g., background knowledge) to facilitate reading. Similarly, in Eskey’s model (1988, p. 96), “*interactive* refers to the interaction between information obtained by means of bottom-up decoding and information provided by means of top-down analysis both of which depend on certain kinds of prior knowledge and certain kinds of information processing skills.” The interactive model assumes that the bottom-up and the top-down processes are equally important.

Readers with different reading styles may prefer one mode of processing over the other. Kirby (1988) classified reading (cognitive) styles into three types—analytic, global, and synthetic. Analytic readers adopt the bottom-up process approach to reading in which the lower-level decoding skills are utilized (e.g., decoding the words). On the other hand, global readers adopt the top-down process approach to

reading in which higher-level cognitive skill is of importance (e.g., employment of background knowledge). However, Kirby (1988) noted that synthetic readers are more successful readers than analytic and global readers. Synthetic readers, according to Kirby, adopt a synthesized version of the bottom-up and top-down approach, that is, an interactive approach. Readers with preference to one approach over the other may not be effective readers. Effective reading requires an adoption of both top-down and bottom-up processing, i.e., the interactive processing. Carrell and Eisterhold (1983) maintained that bottom-up processing ensured that the readers would be sensitive to information that was new or that did not fit their anticipation or prediction of the content or structure of the text, whereas top-down processing helps the readers to resolve ambiguities or to choose alternative possible interpretations of the text. Viewing reading comprehension as involving an interaction of text-based (bottom-up) processing and knowledge-based (top-down) processing, Carrell (1988b) also noted that two kinds of skill deficiencies, linguistic deficiencies and reading skill deficiencies, may lead to inefficient interaction between these two types of processing in ESL reading. Linguistic competence, which involves skills in decoding the vocabulary and syntactic structures of a text, is required for text-based processing (Carrell, 1988b) and successful reading (Eskey, 1988). In reviewing Spiro's (1978) study, Carrell (1988b) accounted for how reading skill deficiencies affected interaction of the two types of processing in comprehension. Lacking text-based skills (e.g., decoding) or knowledge-based skills (e.g., pragmatic inferencing) may result in reading difficulties and the same skill deficiency may lead to overreliance on either text-based or knowledge-based processing. For example, when confronted with reading skill problems, readers may focus on one particular type of processing or they may escape from the problem by shifting to another type of processing to compensate for the processing. In other words, when readers have difficulties understanding texts

by using top-down processing skills, they may try to use bottom-up processing skills to facilitate understanding of texts. For instance, when readers fail to understand the message of a paragraph from context, they may try to break the paragraph into sentences and analyze each sentence in order to interpret the meaning of the paragraph. Hence, as noted by Carrell (1988b), efficient readers shift from one process to the other whereas readers with lower proficiency level tend to focus more on one particular model of processing. Since efficient and effective reading requires both bottom-up reading (lower-level processing) and top-down reading (higher-level processing) operating interactively, it is vital for teachers to teach students to adopt a synthesized style of reading (global & analytic) and to design reading tasks for them to practice and to become familiar with the interactive reading process.

Reading Comprehension and Reading Skills

Two theories are common for describing reading comprehension: one views reading comprehension as a unitary skill which cannot be broken down into separate sub-skills (e.g., Alderson, 2000; Wallace, 1992) whereas the other views reading comprehension as an aggregate of sub-skills such as finding the main idea or interpreting inferences in a text (e.g., Carver, 1978; Grabe, 1986; Grellet, 1981; Mo, 1987; Nuttall, 1996; Raygor & Raygor, 1985). Though there is inconsistency of the evidence for distinct reading skills, it has not prevented material writers to include some of them in EFL reading materials (Williams, E. & Moran, 1989) and so far there has been no evidence to prove that reading subskills do not exist.

Reading comprehension is such a complex cognitive process that researchers attempt to explain the process by analyzing it into a set of component skills. It has been assumed by researchers, teachers, and test writers that reading can be broken down into “underlying skills components” for the purposes of teaching and testing (Grabe, 1991). Researchers suggested that reading comprehension test items be

written to measure reading skills so that they call for different types of reading (e.g., Cohen, 1994; Mo, 1987). Many different lists, taxonomies and even hierarchies of skills have been drawn up by researchers, yet little consensus concerning the content and terminologies can be found (Williams & Moran, 1989). What's more, there has been a fair amount of confusion regarding the use of the terms "skill" and "strategy" in literature. The distinction between these two terms is not clear and they often appear to be used interchangeably. "In EFL reading material, the tendency is for the term "skill" to be used to the exclusion of 'strategy' and even of 'style' (Williams, E. & Moran, 1989, p. 223)." Urquhart & Weir (1998) made a distinction between "skill" and "strategy." They argued that strategies are "reader-oriented" whereas skills are "text-oriented." Richards & Schmidt (2002) identified reading skills as "abilities required for skillful reading" and reading strategies as "ways of accessing the meaning of texts, which are employed flexibly and selectively in the course of reading." In principle, skills are what the reader automatically and unconsciously contributes to the text (text-oriented) while strategies represent conscious decisions taken by the reader to respond to problems during the reading process (e.g., failure to understand a word or sentence) (Williams, & Moran, 1989; Urquhart & Weir, 1998). The present study sees reading as consisting of an aggregate of subskills rather than strategies which readers use to check, monitor, and evaluate their thinking and performance. Detailed discussion of what distinguishes a 'skill' from a 'strategy' is outside the scope of this study.

Theoretical Reading Skill Frameworks

Second/foreign language reading ability was viewed as the mastery of specific reading subskills and this view has continued to inform approaches to the teaching of reading today (Richards, 2002). The investigations of ESL/EFL reading skills have produced a wide variety of skill inventories and classification schemes. Although

there are differences among various classification schemes, most schemes contain one or more of the reading skills listed below, even when the term “strategy” is used rather than “skill” (e.g., Abbott, 2007; Alderson, 1991; Brown, 2003; Grabe, 1986; Grellet, 1981; Mo, 1987; Munby, 1978; Nuttall, 1996; Phakiti, 2003; Purpura, 1997; Raygor & Raygor, 1985; Richards, 2002).

1. Deducing the meaning and use of unfamiliar words
2. Noticing specific details
3. Recognizing cohesive devices
4. Recognizing discourse markers
5. Identifying the main idea
6. Identifying the writer’s point of view
7. Recognizing the organization of text
8. Drawing inferences and conclusion
9. Determining functional values
10. Making predictions

As aforementioned, two researchers in Taiwan (Lu, 2002; Hsu, 2005) adopted Mo’s taxonomy to categorize the reading comprehension items on the SAET and DRET. Mo summarized six types of reading skills by reviewing the categories proposed by Carver (1878) and Raygor & Raygor (1985). Mo (1987) excluded skills unrelated to text structure such as skills related to reading speed, habit and pleasure, and identified six reading skills: (1) identifying the main idea, (2) finding specific details mentioned in the passage, (3) finding implications and drawing inferences and conclusions from the text, (4) recognizing style and tone, (5) determining the special techniques used by the author to achieve his effect, and (6) determining the meaning of strange words or phrases as used in the text. However, as briefly mentioned in Chapter One, Mo’s categorization is so general that it could possibly make it difficult for raters to categorize items appropriately. According to Mo, the fifth skill refers to the skill to understand the organization of a text, the skill to interpret the cohesive

devices used by the writer and any other skills relevant to interpreting the main idea of an article. Thus, the fifth skill in fact involves various different skills. According to Mo (1978), items that measure the reader's ability to recognize cohesive devices, items that measure the reader's ability to recognize the organization of a passage, or items that measure any other techniques that the writer uses to develop the main idea of an article should all be categorized as one skill. This could possibly result in difficulty of item categorization and make the results of categorization unclear. Usually, in most taxonomies, recognizing cohesive devices, which deals with more local information in texts, and recognizing the organization of texts, which deals with more global information in texts, are treated as separate skills. Nevertheless, Mo combined these two skills and other skills into one skill, making the skill ill-defined and complicated. This could possibly be the reason why the researchers in two previous studies, Lu (2002) and Hsu (2005), categorized items that measured cohesive devices and text organization into other types of skills such as items on details or items on inference. Another weakness with Mo's taxonomy was that items which measure test takers' ability to recognize the functional values of sentences or paragraphs were categorized as items on inference since Mo's taxonomy was not able to reflect this skill. Thus, being too general and ill-defined, Mo's taxonomy seemed ineffective in classifying several types of reading comprehension test items on the SAET and DRET.

The Original Nuttall's Taxonomy

Nuttall (1996), in a book designed to teach ESL/EFL reading, developed a reading skill taxonomy which included two major skill categories: word-attack skills and text-attack skills. In terms of word-attack skills, Nuttall (1996) identified three skills required for coping with difficult words while reading: (1) interpreting structural clues, (2) inference from text, and (3) using a dictionary. Knowing what a

word means is recognized as a necessary skill for reading comprehension by researchers. Text attack skills are divided into skills necessary to read for plain sense (bottom-up skills) and skills necessary to understand discourse (top-down skills). Skills to reading for plain sense involve: (1) understanding syntax, (2) recognizing and interpreting cohesive devices, and (3) interpreting discourse markers. Skills necessary to understand discourse are: (1) recognizing functional value, (2) recognizing text organization, (3) recognizing the presuppositions underlying the text, (4) recognizing implications and making inferences, and (5) prediction. A detailed description of Nuttall's taxonomy is presented as follows and modifications made to the taxonomy will be presented in Chapter Three.

Word Attack Skills. Word attack skills are applied when readers need to figure out the meaning of unknown words or phrases. Three word attack skills were identified by Nuttall: interpreting structural clues, inference from context, and using a dictionary.

Word Attack Skill 1: Interpreting Structural Clues

This skill is applied to facilitate understanding when readers are required to determine words by interpreting either of the two following aspects of a word: (1) the grammatical function of the word, and (2) the morphology of the word. By looking at the position of a word, namely its place in the sentence, we can establish its grammatical category and this tells us the kind of meaning to look for. Nuttall also maintained that by adopting an analytic approach to morphology, that is, analyzing the bases, prefixes and affixes of words, readers are able to work out the meanings of new words.

Word Attack Skill 2: Inference from Context

This skill is used when readers try to look for contextual clues in the text in order to infer the meaning of unfamiliar lexical items. According to Nuttall, readers get a

rough idea of the meaning of a word from the context in which it occurred; and with all of the subsequent occurrences of the word, the meaning became more precise to the readers. Nuttall pointed out that this skill is invaluable for less fluent readers.

Word Attack Skill 3: Using a Dictionary

This skill requires readers to look up unfamiliar words in a dictionary. By this skill, readers have to be trained what words should be looked up and what should be done to make use of the information in the dictionary. For example, some readers may find it difficult to find the words they want to quickly enough. Therefore, as Nuttall stated, exercises on alphabetical order and using the guide words at the end of each page will help.

Text Attack Skills. There are a total of eight types of text attack skills: Understanding Syntax, Recognizing and Interpreting Cohesive Devices, Interpreting Discourse Markers, Recognizing Functional Value, Recognizing Text Organization, Recognizing the Presuppositions Underlying the Text, Recognizing Implications and Making Inferences, and Prediction. These skills will be discussed in details in this section.

Text Attack Skill 1: Understanding Syntax

Texts with long sentences or difficult syntax could be difficult to understand even when vocabulary is familiar to the readers. This skill is employed when readers are required to understand long sentences or sentences of difficult syntax. Therefore, students need to break down long sentences into smaller units.

Text Attack Skill 2: Recognizing and Interpreting Cohesive Devices

This skill is applied when readers are required to recognize and interpret cohesive devices in order to facilitate understanding of texts. Items that measure the readers' ability to interpret pro-forms, elliptical expressions or lexical cohesion are considered as measuring this skill. To identify pro-forms, readers are required to

identify the meanings of words like *it, our, this, those, then, one* (as in *the wrong one*), *so/no* (as in *I think so, it appears not*), and comparatives (*smaller, same, additional, such, other*, etc). Words like these are used by writers to avoid needless repetition. With this skill, readers should be able to identify the person, object, or anything referred to and retrieve the reference from the text. This skill is applied when readers are required to identify either anaphoric reference (looking backwards in the text) or cataphoric reference (look forwards). As for interpreting elliptical expressions, readers are required to retrieve the meaning of something left unsaid but has already been mentioned in the text previously. To interpret lexical cohesion, readers are required to interpret the relationship between a lexical item and other parts of the discourse, e.g., synonyms, hyponyms, metaphors, etc.). In sum, items that measure the readers' ability to interpret pro-forms, elliptical expressions or lexical cohesion are considered as measuring this skill.

Text Attack Skill 3: Interpreting Discourse Markers

This skill is concerned with signaling relationships between different parts of the discourse or between the writer and his or her message. Discourse markers often indicate the functional value of the sentence in which they occur and show the relationship the writer intends between two parts of the text. This skill involves interpreting either one of three types of discourse markers: (1) markers that signal the sequence of events (e.g., *then, first, at once, next, the following day*), (2) markers that signal discourse organization (*in conclusion, that is to say, for example, to resume, in short*), and (3) markers that signal the writer's point of view (e.g., *moreover, incidentally, similarly, however, as a matter of fact, in any case, therefore, in order to, if, certainly, more importantly*).

Text Attack Skill 4: Recognizing Functional Value

Nuttall (1996, p. 100) stated that understanding discourse depends on

understanding the functional value of the text sentences. Nuttall divided functional value into three categories: (1) independent functions (associated with propositional meaning; e.g., defining, classifying, generalizing, naming, describing, reporting, speculating, predicting, etc.), (2) text-dependent functions (associated with contextual meaning; e.g., asserting, exemplifying, explaining, reinforcing, explicating, hypothesizing, commenting, concluding, etc.), and (3) interaction-dependent functions (associated with pragmatic meaning; e.g., conceding, evaluating, inviting, instructing, apologizing, suggesting, complaining, complimenting, warning, etc.).

Text Attack Skill 5: Recognizing Text Organization

Skill of this type is used when the reader is required to trace the rhetorical development of a text and to analyze the rhetorical organization of a text. In general, paragraphs can be organized in many patterns (e.g., definition, description, cause-effect, classification, comparison-contrast, process, hypothesis, argumentation, exemplification, etc.). It is easier for readers to interpret difficult sentences if they can identify the principle by which the text is organized and see how the ideas hang together (Nuttall, 1996, p.106).

Text Attack Skill 6: Recognizing the Presuppositions Underlying the Text

This skill is applied when the reader is required to figure out presuppositions such as the knowledge and experience that the writer expects the reader to have or the opinions, attitudes, emotions that the writer expects the reader to share, or at least to understand.

Text Attack Skill 7: Recognizing Implications and Making Inferences

This skill is necessary when the reader is expected to understand information implicitly stated in the text or draw certain unstated conclusions, points in an argument from the text. The answer of a question of this type cannot be directly located but has to be inferred.

Text Attack Skill 8: Prediction

This skill involves readers using their understanding of the passage and their background knowledge to determine what might happen next or after a text ends. Prediction involves using schemata about the way stories work, the way texts are constructed, and the way people tend to think. By prediction, readers are able to predict what the writer is likely to put forward next.

In sum, Nuttall's taxonomy consists of three word attack skills and eight text attack skills. The structure of Nuttall's taxonomy is presented in Table 1 below.

Table 1. Structure of Nuttall's Original Taxonomy

Word attack skills
1. Structural Clues
2. Inference from Context
3. Using a Dictionary
Text attack skills
Skills necessary to read for plain sense (Bottom-up skills)
1. Understanding Syntax
2. Recognizing and Interpreting Cohesive Devices
3. Interpreting Discourse Markers
Skills necessary for reading beyond plain sense and into discourse (Top-down skills)
4. Recognizing Functional Value
5. Recognizing Text Organization
6. Recognizing Presuppositions Underlying the Text
7. Recognizing Implications and Make Inferences
8. Prediction

Note. Adopted from Nuttall's original taxonomy: *Teaching Reading Skills in a Foreign Language*, by Nuttall, 1996.

However, a few weaknesses and practical limitations could be found in the original taxonomy. The skills proposed by Nuttall are important for efficient and effective reading; however, skills which are not supposed to be measured on the reading comprehension section and skills which are not able to be measured in testing

situations should be excluded for the purpose of the present study. In terms of word attack skills, the first word attack skill “Interpreting Structural Clues” is usually measured in the section which tests vocabulary knowledge, not in the reading comprehension section. Therefore, this skill is excluded from the analysis of the current study. In addition, the second word attack skill “Using a Dictionary” is obviously not applicable in testing situations as the examinees are not allowed to use a dictionary during testing. Therefore, this skill is also excluded. Thus, the only skill left is the second word attack skill “Inference from Context.” As for the skills under text attack skills, some skills are not measured or are usually measured in other sections on the SAET or DRET rather than in the reading comprehension section. First, “Understanding Syntax” requires the reader to break down long sentences in order to arrive at meaning. According to Nuttall, when a reader encounters structural difficulty, he or she must take an analytic approach to analyze long or difficult sentences. For example, the reader is suggested to identify the cohesive elements, rewrite the sentence, find the nouns, and identify the verbs, etc. These are more related to analyze the grammar rules of a sentence, but reading comprehension tests do not require an examinee to analyze the grammar of sentences. Therefore, this skill is excluded as well. Secondly, “Interpreting Discourse Markers” which requires the reader to identify and interpret discourse markers is usually measured in the cloze test instead of the reading comprehension test. Thirdly, “Prediction” involves readers using their understanding of the passage and their background knowledge to determine what might happen next or after a text ends. In testing situations, this skill cannot be used and thus is excluded as well. Thus, the original taxonomy was revised to fit the purpose of the present study and the revised taxonomy is presented in Chapter Three.

Reading Skill Instruction

With strengthened English reading skills, ESL/EFL readers will make greater progress and attain greater development in all academic areas of English (Anderson, 1999). As many researchers have assumed that reading skills can be analyzed as involving subskills, the way to help learners improve their reading is to practice those subskills (Dubin, Eskey & Grabe, 1986; Grellet, 1981; Mo, 1987; Nuttall, 1996). Researchers have stressed the importance of teaching reading skills to facilitate reading comprehension and to achieve successful reading by teaching one or many reading skills in various existing taxonomies (e.g., Duke & Pearson, 2002; Dunmore, 1989; Mo, 1987; Nuttall, 1996; Pressley, 2000; Rubin, 1982; Su, 2001; Williams, 1983). Mo (1987) pointed out that English reading instruction in Taiwan placed little emphasis on the comprehension of textual structure; therefore, students often understand every sentence of a text yet fail to understand textual organization and the main idea of a paragraph. Therefore, he proposed that reading skills should be taught to help students understand textual organization and to improve their reading abilities. He also proposed that reading tests be constructed by incorporating reading skills into test items.

Empirical studies have showed that explicit reading skill instruction enhances EFL reading (e.g., Chen, 2005; Su, 2001). Chen (2005) investigated the effects of explicit instruction of reading skills on EFL senior high school students' reading comprehension and also examined the students' comprehension of different types of questions. What's more, the students' perceptions of reading skills, reading attitudes, and their responses to the skill instruction were investigated. A total of 89 year-three students from two classes of a senior high school in Kaohsiung were recruited as participants in Chen's study. A three-month explicit skill instruction was provided with a focus on five skills: (1) skimming for main ideas, (2) identifying topics and

main ideas, (3) making predictions, (4) making inferences, and (5) guessing the meanings of unfamiliar words from context. These skills were taught in an explicit approach which emphasized modeling, guided practice, independent practice, and application. Furthermore, reading comprehension tests and questionnaires were conducted before and after the treatment to collect the related data for analysis and discussion. The results showed that the skill instruction proved effective in building up the participants' knowledge and use of the skills. According to Chen, most of the participants did not have a clear understanding of those skills before the treatment; however, they had learned to use them and liked to use them after the treatment. The results also revealed that the skill instruction helped enhance the participants' learning interests toward English reading and their confidence in their English reading ability. Also, it was found that the effect of skill instruction was more significant on the intermediate and the lower proficiency participants in promoting reading comprehension. The lower proficiency subjects benefited most from the treatment and most participants had a positive response toward the skill instruction. In terms of the effectiveness on each reading skill, the skill instruction helped improve the participants' comprehension for main idea, detail, and inference questions, but not for word-guessing questions. Among the skills instructed, "guessing the meanings of unfamiliar words from context" and "skimming for the main idea" were considered the most practical skills by the participants.

Su (2001) conducted an empirical study to evaluate effectiveness of teaching English reading skills and the cognition of the students on the reading skills they learned. Su introduced fourteen reading skills to her students in the first-year college English reading course whenever her students need to apply those skills in their reading. The reading skills Su introduced to the participants were (1) questioning, (2) previewing and predicting, (3) scanning, (4) recognizing topics, (5) resolving

ambiguities in the text, (6) looking for key words, (7) guessing vocabulary from context, (8) locating topic sentences, (9) understanding main idea & details, (10) skimming, (11) reading aloud, (12) small group discussion, (13) paraphrasing, and (14) role-playing. According to Su, these skills were selected for students who need “elementary” literacy skills to help them read and think in English. After the reading skills were introduced to the students, Su evaluated the students’ feelings about the reading skills. A questionnaire was designed for the participants to fill out at the first and the final classes of the second semester of instruction to rate the usefulness of the reading skills they learned. The results showed that all the students in Su’s study agreed that the reading skills they learned were helpful for the improvement of reading ability except for one skill—the skimming skill. In addition, most of the students had difficulties with reading skills such as skimming, paraphrasing and guessing vocabulary from context.

Although researchers have different opinions regarding how many reading skills there are, in general, it is agreed that the teaching of reading skills is helpful in promoting reading comprehension. Also, it is very important for teachers to be aware that if reading skills are teachable, then they are probably testable (Nuttall, 1996). Nuttall (1996, p. 220) argued that “If we do no test what we are trying to teach, there is a danger that students will not pay much attention to what is taught, or that teachers will cease to teach it.”

Testing ESL/EFL Reading Comprehension

In this section, the relationship between teaching and testing is discussed. Also, literature related to using multiple choice questions to measure reading skills is presented.

Why Testing (The Relationship between Teaching and Testing)

Testing is an important part of every teaching and learning experience (Madsen, 1983). The proper relationship between teaching and testing is surely that of partnership (Hughes, 2003). Testing should not only follow teaching but should support good teaching and exerts a corrective influence on bad teaching. Therefore, understanding of tests could provide information about whether our teaching needs improvement or certain part of our teaching needs more attention. The importance of testing has been emphasized by a number of researchers. As McNamara (2000) pointed out, language tests can serve as a valuable tool for teachers by providing information that are relevant to language learning: such as (1) evidence of the results of learning and instruction as well as feedback on the effectiveness of the teaching program, (2) information relevant to decision-making about individuals, and (3) information for teachers to clarify instructional objectives. Tests can provide information about teaching and learning and thus tests can be helpful to both teachers and students.

According to Madsen (1983), well-made tests can help students develop positive attitudes toward instruction by giving them a sense of accomplishment and a feeling that the teacher's evaluation of them corresponds to what has been taught. In addition, tests also help foster learning since they help the teachers confirm which part of the learning each student has mastered and which part needs further attention and improvement. What's more, good English tests can assist the students in learning the language by requiring them to study hard, emphasizing course objectives, and showing them where they need to improve.

The qualities desirable for a good test are validity, reliability, and practicality (Bachman, 1990; Brown, 2001; Harries, 1969; Hughes, 2003). These qualities are generally regarded as the basic requirements of a good test. Validity is considered as

the most complex criterion of a good test (Brown, 2003). A test is said to have validity when it actually measures what it intends to measure. For example, if the purpose of a reading test is to examine the reading ability of a particular group of students, the test results should reflect their true reading ability. Validity is an important factor in designing good reading comprehension tests (Sequera, 1995). Three types of important validation of a test are content validity, face validity, and construct validity. The core element of test validity is the construct—the theoretical representation of the skill or knowledge that the test purports to measure (Slomp, 2005, p. 149). As such, the validity of a test rises or falls in accordance with the degree to which its scores are a reflection of students’ ability in relation to the construct. Hughes (2003) defined construct in language testing as “any underlying ability (or trait) that is hypothesized in a theory of language ability.” (p. 31). Hence, in the case of reading comprehension, tests should reflect the theoretical assumptions under which reading teachers operate (Sequera, 1995). Therefore, it is assumed that the tests reflect the objectives of reading abilities stipulated in the curriculum guidelines and it is one of the issues that the present study aims to address.

Testing Comprehension with Multiple-choice Questions

Many textbooks on language testing (e.g., Heaton, 1988; Hughes, 1989; Weir, 1990, 1993) have given examples of testing techniques that might be used to assess language abilities. Among various test techniques, multiple-choice (hence MC) format is a common device for testing students’ reading comprehension. Multiple-choice (MC) is considered to be an objective technique which “requires intellectual discrimination skills, a versatile test capable of probing a variety of areas and different types of cognitive activities such as acquisition of knowledge, understanding, application, analysis and evaluation” (Green, 1975; Marshall, 1971, cited in *Nevo*, 1989). The most obvious advantage of MC questions is probably in that

the scoring can be reliable, rapid, and economical. In addition, MC format allows for more items to be included in a given period of time. A further advantage is that MC format allows the testing of receptive skills without requiring the test-takers to produce output.

Nevertheless, given the virtues of multiple-choice questions, this format has been criticized by researchers on a number of grounds. A serious disadvantage with MC questions is that guessing may have a considerable but unknowable effect on test scores (Hughes, 2003). Test-takers may get an item correct for the wrong reason without actually understanding the text. Another objection to the use of multiple-choice questions is that they are often passage-independent which means that the items could be answered without reference to the reading passages accompanying them (Bernhart, 1991; Teale & Rowley, 1984; Weir, 1997). Evidence has been found that test-takers do not always recourse to the reading passages in response to MC questions (Cohen, 1998; Nevo, 1989); it is suspected that explicit teaching of certain techniques will help students become test-wise and thus improve their scores (Richards, 1997). Nevo (1989, p. 212) argued that “it would appear useful to devote attention, time, and effort to guiding and training students in coping effectively with a test format like [the multiple-choice test]”.

A further concern with the use of multiple-choice (MC) format is that there has been much doubt with respect to the validity of multiple-choice (MC) questions. Weir (1990) and Urquhart & Weir (1998) argued that answering multiple-choice questions is an unreal task since in real life communications one is rarely required to choose one answer from four choices to show understanding. In a multiple-choice test, distractors can be used to trick the test-takers by presenting choices that the test-takers may not have otherwise have thought of (Alderson, 2000; Richards, 2000; Weir, 1990). Test-takers could be deliberately tricked into confusing dilemmas. Therefore, it is

difficult to know whether failure of questions is due to lack of comprehension of the text or lack of comprehension of the question.

Despite of the criticisms against the use of the MC technique, multiple-choice questions has been widely used to assess reading ability and therefore issues have been raised about what MC questions actually measure in reading tests and whether they are valid measurements of reading ability (Cummings, 1982; Farr, Prichard & Smitten, 1990; Urquhart & Weir, 1998; Weir, 1997). Weir (1997) proposed a four-level version of reading comprehension for testing purposes: reading expeditiously for global comprehension, reading expeditiously for local comprehension, reading carefully for global comprehension, and reading carefully for local comprehension. In Weir's version of operations in reading, skills such as understanding the syntactic structure of a sentence and clause, understanding lexical and/or grammatical cohesion, understanding word meaning and locating specific details are bottom-up skills operated when reading at local levels, while identifying the main idea and making inferences are top-down skills operated at global levels. Similarly, Harrison (1983, in *Navo*, 1989) argued that through the MC format, it is considered possible to check all reading levels (the semantic and syntactic aspects of the test), the discourse level (cohesion and coherence connections among various parts of the text), and the pragmatic level (the writer's point of view). In other words, bottom-up reading skills and top-down reading skills could be tested via MC questions.

Studies of Reading Tests Analysis (Item Analysis)

As noted previously, researchers agree that tests can provide information about teaching and learning and thus tests can be helpful to both teachers and students. An analysis of the reading comprehension test items also reveals the knowledge and skills valued by the reading comprehension tests. Research into reading tests has received

much interest in the field of reading research in Taiwan. Both SAET and DRET exert considerable influence on English teaching and learning in Taiwan because the scores a student receives on each exam is a determining criterion for admission to desired universities. Thus, with the importance of both tests, they have been topics for discussion and research reports on the overall test content analysis and statistical analysis of both tests have been conducted annually by the College Entrance Examination Center (CEEC Web site; Huang, 1994; Jeng 1992; Jiang & Lin, 1999; Xu & Lu, 1998). In particular, most of the studies on the SAET and DRET aimed at providing an overview of the test construction or statistical results of difficulties of items, distractors, and passing rates, etc. rather than a thorough qualitative analysis on the reading skills tested on each comprehension item in particular (e.g., Huang, 1994; Jeng, 1992; Jiang & Lin, 1999; Xu & Lu, 1998). Most statistical analyses have often focused on the numbers of items and distribution, length of text, vocabulary, topics, discriminatory power, and examinees' test performances (e.g., passing rates). Huang (1994) conducted a qualitative analysis of the Joint College Entrance Examination (hence JCEE, renamed as DRT in 2002) English test items from 1985 to 1994. The results of the reading comprehension item analysis showed that roughly over 90% of the items were well-designed, yet a few were not well-constructed or not even reading comprehension items at all. Items designed to test examinees' vocabulary in context are reading comprehension questions on account that the readers have to look for contextual clues in the texts. However, Huang found that some items designed to test examinees' vocabulary were not well-written because they could be answered without referring back to the text. What's more, he found that some items were designed to test examinee's grammar knowledge instead of their reading abilities. He indicated items designed to test examinees' knowledge of vocabulary and grammar without referring back to the text are ill-written and thus should be excluded. He also said that

cautions should be taken when constructing items that involve arithmetic.

Jeng (1992) conducted a statistical analysis of the English test of JCEE in 1991. This analysis focused on the difficulty of items, discriminatory power, and distractors in the test. There were three formats of questions, including sections on conversations, cloze, and reading comprehension. It was found that 76% of all of the items in the test, a total of 34 items, were well-written. Among the well-written items, 13 items were reading comprehension items, which amounted to 86% of the well-written items in total. The results also showed that the overall reliability of the test items reached .90. Xu and Lu (1998) studied various elements of the JCEE English test content in 1998, including the topics, text length, syntactic complexity, vocabulary, distractors, and question types. The researchers found that these reading comprehension test items could usually be classified into four types: vocabulary, main idea, detail, and inference. However, they did not further identify those elements item by item or examine the frequency and distribution of different items. The aforementioned studies aimed at exploring the overall test construction or statistical results rather than a thorough qualitative analysis on the reading skills tested on each comprehension item in particular.

Recently, two studies (Hsu, 2005 ; Lu, 2002) used Mo's taxonomy as the coding scheme to analyze the reading comprehension test items on SAET and DRET. Lu (2002) conducted both qualitative and quantitative analyses of the reading skills measured in reading comprehension section on the SAET from 1995 to 2002. Qualitatively, Lu categorized the reading comprehension test items by using Mo's classification as the coding scheme. Lu also examined the texts and the variables that affected the passing rates of high achievers and low achievers. With respect to quantitative analysis, she computed the frequency distribution of question types and the correlation between question types and passing rates. Results showed that the

most frequent question types were the items on details, followed by items on inference, main idea, writer's style/tone, organization, and word meaning. Lu further categorized the items on detail into seven types: (1) specific-answer questions, (2) identifying true/false statements, (3) cause-effect, (4) number/date, (5) contrast, (6) sequence-of-events, and (7) following-directions. Lu's findings also revealed that in general, all of the examinees performed best on items that test word meaning, followed by items on main idea, details, and inference, while they performed worst on items that test organization and style/tone. With respect to the performances by high and low achievers, it was found that high achievers performed best on items that tested details, followed by items on inference, word meaning, and main idea whereas low achievers on average performed fairly on items that tested word meaning and main idea but badly on items that tested details and inference. Neither group of achievers performed well on items that test the writer's style/tone. Both high and low achievers have difficulties in several areas such as answering items that demand for higher-order reading processes, answering items that require interpretation and inference, synthesizing numerous details to get the item right, recognizing textual features, understanding lengthy articles with unfamiliar topics, and so on. However, I found that that Lu's study exhibited several weaknesses. First, the analysis in Lu's study is subject to the researcher's own interpretations. It would have been better if another rater is recruited to analyze the test items to see if there is consistency between raters. Secondly, it was found that the analysis of items on details and on inference is confusing because Lu categorized some items which tested inference skills into different categories, some as items on details and others as items on inference of details. Thus, it is possible that Lu's method of categorization is flawed.

Hsu (2005), applying the same coding scheme, attempted to analyze the reading comprehension test items on the 2000 & 2001 JCEE English test and DRET from

2002 to 2004. The themes of the texts, text variables that accounted for item difficulty, the examinees' passing rates on each question type and the discrimination index of the test items were also investigated. Moreover, the Word List published in 1996 by the CEEC was used to analyze the use of words in the chosen texts to explore which words were beyond the scope of the Word List. To examine the passing rates on each question type, instead of computing the statistical data of passing rates provided by the CEEC, Hsu recruited 76 year-two students (divided into the high-proficiency group of 20 students, the middle-proficiency group of 36 students, and the low-proficiency group of 20 students) from two classes in a high school in Kaohsiung city as participants to answer questions based on eighteen passages on the 2001 JCEE and 2002-2004 DRET and used their performance for data analysis. Similar to Lu's (2002) study, it was found that the most frequently used type of question is the detail item. Likewise, examinees performed well on items that tested lower-level skills such as determining the meaning of words and finding specific details, whereas they performed worst on questions of drawing inferences, which required higher level processing. As for the passing rates of high and low achievers, it was found that among the four most frequent types of questions (items on word meaning, detail, main idea, and inference), the high-proficiency group performed best on items that tested details and word meaning yet worst on items that tested main idea. The low proficiency group performed best on the word meaning item yet they performed worst when requested to draw inferences and identify the main idea of a text. Both high and low group performed best on the detail item. Similar to Lu's (2002) study, the findings revealed that items on details were the most frequently tested. Likewise, examinees performed well on items that tested lower-order skills such as determining the meaning of words and finding specific details, whereas they performed worst on items that tested higher-order skill such as drawing inferences. However, several

drawbacks can be found in Hsu's study. Firstly, Hsu's study would have been more convincing if the author had directly computed the passing rates of examinees actually taking those tests rather than recruiting students who did not take the test as participants. If Hsu had computed the passing rates of the examinees who actually took the test, the results may turn out differently. Secondly, it was found that the categorization of some test items by Hsu is problematic. It has been found that items that measure test takers' knowledge of discourse were categorized as items testing word meaning and items testing details in Hsu's study. Take the item analysis of DRET 2002 for example, Q 56 was categorized as a word meaning item by Hsu while it in fact tested discourse knowledge of referent "*he*." An examinee is required to look for contextual clues in order to assign meaning to the referent *he*. Q 51 on the 2004 DRET, which tested discourse knowledge of referents as well, was classified as a detail item in Hsu's study. Thus, it is assumed that all items that tested discourse knowledge were completely categorized incorrectly as word meaning items and detail items by Hsu. Therefore, the results of Hsu's study may be problematic.

A more recent study by Lan (2007) aimed to analyze the reading comprehension question types on the SAET and DRET both qualitatively and quantitatively by adopting a revised Bloom's taxonomy. Lan's study aimed to investigate what cognitive process levels and knowledge types were tested in both SAET and DRET from 2002 to 2006. The revised Bloom's taxonomy identified by Anderson and Krathwohl (2001) was divided into two dimensions—the knowledge dimension and the cognitive process dimension. The knowledge categories of the revised Bloom's taxonomy contains four major types of knowledge with subtypes under each category—factual knowledge, conceptual knowledge, procedural knowledge, and metacognitive knowledge. The cognitive process dimension includes six cognitive process categories and each includes subcategories. The six major types of cognitive

process dimensions are *Remember, Understand, Apply, Analyze, Evaluate and Create*.

The results of Lan's study showed that from the item classification of the 2002 to 2006 SAET and DRET, four major cognitive process levels along with eight sub-levels and three types of knowledge along with four subtypes were found. The four main cognitive processes identified were: *Remember, Understand, Apply, and Analyze*. Five major combinations of cognitive levels and types of knowledge found in the study were: (1) Remember Factual Knowledge, (2) Understand Factual Knowledge, (3) Understand Conceptual Knowledge, (4) Apply Procedural Knowledge, and (5) Analyze Conceptual Knowledge. Nine sub combinations were identified as well: (1) Recognizing specific details and elements, (2) Interpreting specific details and elements, (3) Inferring specific details and elements, (4) Classifying into classifications and categories, (5) Summarizing principles and generalizations, (6) Inferring classifications and categories, (7) Explaining principles and generalizations, (8) Executing subject specific skills and algorithms, and (9) Attributing principles and generalizations.

The results of Lan's study revealed that in 2002 to 2006 SAET and DRET reading comprehension test item analysis, 5 major question types were identified: (1) Remember Factual Knowledge, (2) Understand Factual Knowledge, (3) Understand Conceptual Knowledge, (4) Apply Procedural Knowledge, and (5) Analyze Conceptual Knowledge. Among the question types emerged, items on Remember and Understand Factual Knowledge were the most frequently tested throughout the five years, accounting for 74.1% in the SAET and 73% in the DRET. Items at the Evaluate and Create levels were not found.

The result of item frequency in the SAET showed that around half of the items aimed to test students' ability to recognize facts in the passages and almost one third of the items measured the ability to understand specific details. Similar to the SAET,

the DRET focused on testing students' abilities to identify and to understand facts as well. In terms of frequency, both tests showed a similar pattern since most of the items were Remember Factual Knowledge, followed by Understand Factual Knowledge, Understand Conceptual Knowledge, Analyze Conceptual Knowledge, and Apply Procedural Knowledge. For the frequency of the major questions types, two types of items—items on Remember Factual Knowledge and Understand Factual or Conceptual Knowledge—were the majority in both SAET and DRET throughout 2002 to 2006 (accounted for 85.8% on the SAET and 88.9% on the DRET). Few items were found at higher cognitive skills of Apply and Analyze (accounted for 14.3% on the SAET and 11.1% on the DRET).

Comparing SAET with DRET, it was found that the major differences were the frequency, occurrence, and distribution of the cognitive sub-skills and knowledge types (Lan, 2007). In terms of frequency, the SAET had more *Recognizing* items than the DRET. In terms of occurrence, the study showed that the DRET consistently covered various items subtypes each year whereas the SAET fluctuated in its variety of item types. The SAET consisted of at least three subtypes each year while the DRET had five question subtypes on average each year, except 2003 when six question types emerged. Lan (2007) thus concluded that the results showed that the cognitive skills and knowledge types tested in both tests were relatively fixed.

The results of Lan's study (2007) showed that five major item types accompanying nine subtypes emerged in the SAET and DRET. Recognizing items (known as Remember items) and Interpreting items (subtypes of Understand Factual Knowledge items) were the majority. It was also found that both tests seemed to emphasize measuring test-takers' abilities to identify and understand specific elements or details in the reading passages while only a few items asked the test-takers to understand the reading passages as a whole through summarizing,

classifying, establishing a cause-and effect model, or inferring. What's more, it was found that Executing questions that involved mathematical operation tend to have more occurrences in the SAET, whereas more inferring items on inferential facts or concepts were common in the DRET. On average, the overall passing rates on all items in the DRET ($M=44.62$, $SD=12.83$) were lower than those in the SAET ($M=56.42$, $SD=14.19$) from 2002 to 2006.

The present study, different from previous research on test item analysis by using Mo's taxonomy or the revised Bloom's taxonomy as the coding scheme or other methods, attempts to employ a revised Nuttall's taxonomy to analyze the reading skills operating in the reading comprehension test items on both SAET and DRET from 2002 to 2007.

Summary

This chapter has so far explored the rationale behind the reading processes, reading skill frameworks, the reading skill taxonomy by Nuttall, studies on the instruction of reading skills, and the studies on reading comprehension tests. Five sections are included in this chapter. In the first section, a brief discussion on the reading process models was presented, introducing three major types of reading models—the bottom-up model, the top-down model, and the interactive model. Efficient and effective reading requires both bottom-up reading and top-down reading operating interactively. In the second section, issues of reading comprehension skills were discussed and reading skill frameworks were reviewed. It has been assumed by researchers, teachers, and test writers that reading can be broken down into “underlying skills components” for the purposes of teaching and testing (Grabe, 1991). In addition, many researchers have assumed that the only way to enhance reading comprehension is to practice reading skills (e.g., Dubin, Eskey & Grabe, 1986; Grellet, 1981; Mo, 1987; Nuttall, 1996). The third section presented the original Nuttall's

taxonomy in details. However, Nutall's taxonomy will be revised for the purpose of the present study and the revision will be presented in Chapter Three. In the section on reading skill instruction, empirical evidence was provided to prove that reading skill instruction helps to improve students' reading ability. In the last section, issues related to testing reading skills were explored and problems and limitations of previous studies which explored the reading skills in SAET and DRET were presented. Lu and Hsu adopted Mo's taxonomy as the coding scheme to investigate the reading skills measured in the SAET and DRET separately while Lan used the Revised Bloom's Taxonomy as the coding scheme to examine the cognitive levels and knowledge types measured in both SAET and DRET. In general, the results in these three studies showed that items on details are the most frequently measured. However, weaknesses have been found with Lu's and Hsu's study. In addition, although the taxonomy adopted by Lan is well-designed and could be applicable in many fields of research, there is little research done by using a taxonomy designed specifically for ESL/EFL reading skills to analyze the reading comprehension test items on both SAET and DRET. Thus, it is mainly to fill this gap that the present study is conducted.

CHAPTER THREE

METHOD

This chapter describes the method employed in the present study. This includes the reading materials, the instruments for the qualitative and quantitative analysis, and the procedure of data analyses.

Materials

Data in the present study, including the reading materials, test items, passing rates⁴ of all examinees and those of high achievers and low achievers⁵, was collected from the College Entrance Examination Center (CEEC). The reading materials used in this study consisted of reading passages in the Scholastic Achievement English Test (SAET) and Department Required English Test (DRET) from the year 2002 to 2007. The reading materials from the year 2002 to 2007 are chosen on account that a new education policy was introduced in 2002 when the traditional JCEE was abolished and the Scholastic Achievement Test (SAT) and Department Required Test (DRT) became the two entrance exams for entering universities. A total of 43 passages, 24 on the SAET and 19 on the DRET, and 167 reading comprehension test questions, 93 on the SAET and 74 on the DRET, were collected. In general, each reading comprehension test contains three to four passages and each passage was followed by three to five questions relating to its content. All of the reading comprehension test items were analyzed and categorized according to the reading skills measured in each item.

Table 2 presents the number of reading passages and their comprehension test

⁴ The passing rate, also called item facility, shows the proportion of examinees that got an item correct. It is suggested that acceptable range of item facility is 33 to 67 (Alderson et al., 1995; Brown, 1996; Henning, 1987; McNamara, 2000; Yu, 1997). Jeng et al. (1999) suggested that the ideal item facility index for national examination such as college entrance examination should be 30 to 80.

⁵ The high and low achievers refer to the examinees who reach the top and the bottom 33% in the population of the SAET and DRET.

items under analysis.

Table 2. The Number of Reading Passages and Comprehension Test Items

Year	SAET		DRET	
	Number of Reading Passages	Number of Test Items	Number of Reading Passages	Number of Test Items
2002	4	15	4	15
2003	4	15	3	15
2004	4	15	3	11
2005	4	16	3	11
2006	4	16	3	11
2007	4	16	3	11
Total	24	93	19	74

Participants

The information of sampling in both SAET and DRET are shown in Table 3 and Table 4. The statistics were obtained from the CEEC Web site. As shown in Table 3 and Table 4, the total number of examinees in the SAET was around 150,000 while in the DRET the total number of examinees was between 100,000 and 125,000. The sampling size varied each year since the students can choose to take either test. Most students would take the SAET while only those who hope to obtain better scores in the DRET to apply for their preferred universities would take the DRET. Thus, students who had done well on the SAET might not take the DRET.

Table 3. The Number of Examinees Taking the SAET from 2002 to 2007

Year	2002	2003	2004	2005	2006	2007
Number of Examinees	146,638	164,521	157,544	160,522	161,567	153,364

Table 4. The Number of Examinees Taking the DRET from 2002 to 2007

Year	2002	2003	2004	2005	2006	2007
Number of Examinees	113,193	125,608	117,705	116,492	109,770	100,059

Instrument

To answer the research questions of the present study, both qualitative and quantitative data analyses were performed. Nuttall's taxonomy was revised because of some limitations and the revised taxonomy was adopted as the coding scheme for the qualitative analysis of all reading comprehension test items under analysis.

The original Nuttall's taxonomy was presented in Chapter Two. Different from Mo's scheme, Nuttall's taxonomy, which was designed for ESL/EFL reading instruction, can be adopted to analyze items that might be inappropriately classified in Mo's taxonomy. Since it is designed specifically for EFL/ESL reading instruction, Nuttall's taxonomy fits the EFL context of Taiwan. In addition, in Nuttall's taxonomy, the ability to recognize cohesive devices and the ability to recognize text organization are treated as separate skills. Therefore, Nuttall's taxonomy would be a better taxonomy than Mo's in identifying the skill to recognize cohesive devices and the skill to recognize text organization since these two skills were measured in either SAET or DRET. What's more, Nuttall's taxonomy contains skills that were measured either on the SAET or the DRET but items which measure these skills were not identified in Mo's taxonomy. For example, one skill on Nuttall's taxonomy—*Recognizing Functional Value*—had been measured on the DRET (See Appendix 1 for the principle of categorization and example question of this skill); however, in Mo's taxonomy, items which test this skill were classified as items on inference. Thus, we can assume that Nuttall's taxonomy could better identify skills

tested than Mo's taxonomy. Nevertheless, as a senior high school English teacher, I found that Nuttall's taxonomy has a few weaknesses and practical limitations. Therefore, a revised taxonomy is designed by using Nuttall's rubrics as the baseline classification scheme and incorporating other important skills lacking in Nuttall's original taxonomy.

As mentioned previously in Chapter Two, due to some limitations of the original Nuttall's taxonomy, some reading skill types were excluded from the original scheme. Apart from the aforementioned limitations, Nuttall's taxonomy is considered to be lacking several common and important reading skills which can be found in most reading skill taxonomies and these skills should be added to the original taxonomy. While researchers have classified reading skills in somewhat different perspectives, there seems to be substantial agreement on the significance of such skills as identifying the main idea and identifying details (e.g. Abbott, 2007; Brown, 2003; Carver, 1978; Grabe, 1986; Grellet, 1981; Mo, 1987; Phakiti, 2003; Raygor & Raygor, 1985; Williams and Moran, 1989). Nevertheless, these two skills are lacking in Nuttall's original taxonomy. Understanding the main idea is widely recognized as a major goal for reading comprehension. What's more, in reading comprehension tests, readers are often required to locate specific details in the passages. Therefore, these two skills will be incorporated into the original taxonomy so as to form a revised version of Nuttall's taxonomy. In sum, the skills excluded from the original taxonomy include "Structural Clues," "Using a Dictionary," "Understanding Syntax," "Interpreting Discourse Markers," and "Prediction" while the skills added to the original scheme involve understanding the main idea and identifying details. The present study, taking a skill-oriented perspective, attempts to analyze the reading comprehension test items in both SAET and DRET by implementing a revised Nuttall's taxonomy to see what reading skills are measured in both tests. Descriptions

of the skills incorporated into the original taxonomy and the scheme of the revised taxonomy as shown in Table 5 are presented in the next section. Principles for classification of items based on the revised Nuttall's taxonomy with example questions are shown in Appendix 1.

Table 5. Structure of the Revised Nuttall's Taxonomy

Skills necessary to read for plain sense (Bottom-up skills)
1. Word Inference from Context
2. Recognizing and Interpreting Cohesive Devices
*3. Identifying and Interpreting Details
Skills necessary for reading beyond plain sense and into discourse (Top-down skills)
3. Recognizing Functional Value
4. Recognizing Text Organization
5. Recognizing Presuppositions Underlying the Text
6. Recognizing Implications and Making Inferences
*7. Recognizing and Understanding the Main Idea

Note: This taxonomy is developed by revising Nuttall's taxonomy to suit the purpose of the present study. The third and seventh skills are skills incorporated into the original taxonomy.

In sum, to form a framework based on Nuttall's taxonomy with modifications, two skills under word attack skills and three text skills were excluded and two important skills lacking in the original taxonomy were incorporated. Since there is only one skill under word attack skill left—"Inference from Context," the researcher decided to rename it as "Word Inference from Context" in order to distinguish it from "Recognizing Implications and Making Inferences" and put it under the category of "Skills necessary to Read for Plain Sense. "Word Inference from Context" is used when readers are required to look for contextual clues in the text in order to infer the meaning of unfamiliar lexical items whereas "Recognizing Implications and Making Inferences" is necessary when readers are expected to understand information implicitly stated in the text or draw certain unstated conclusions by looking for the

connections between the options (including distracters and the correct answer) in the questions with the information presented in the text. “Word Inference from Context” is inference at word level and thus is classified as a skill under the bottom-up skill. In addition, since some skills were excluded and others were incorporated, the order of skills has been rearranged. Therefore, in the revised Nuttall’s taxonomy (see table 3), there are a total of eight skills.

Data Analysis

The data was analyzed by looking at reading skills measured and the passing rate of each skill. First, reading comprehension test items in both SAET and DRET were categorized by using the revised Nuttall’s taxonomy to find out what reading skills were measured on both tests. Secondly, the frequencies of items that measured different reading skills in both tests were calculated to obtain information about the differences between reading skills measured on the SAET and DRET. Thirdly, the passing rates of all examinees on each item type were analyzed to gain insights into the strengths and weaknesses of students’ reading skills and to draw implications on EFL reading instruction. Fourthly, the passing rates of high achievers and low achievers were analyzed to see how high achievers and low achievers differ with regard to the reading skills measured. The analysis of reading comprehension test items proceeded in three phases: (1) preliminary analysis, (2) formal item analysis, and (3) analysis of the data and the passing rates.

Formal Item Analysis

Apart from the researcher as a rater, two more raters, including an MA student in the TESOL program at National Taiwan Normal University, and an MA graduate in the same program, were recruited to ensure the validity of coding results of the present study. Prior to the coding procedure, the descriptions of the original taxonomy, the taxonomy developed specifically for the present study (the Revised Nuttall’s

taxonomy, see Table 3), the criteria for coding and example questions (Appendix 1), the coding sheets, and the reading comprehension test items on 2002-2007 SAET and DRET were given to the raters to categorize each item.

Before the raters started the analysis of test items, all raters participated in a training session and the researcher discussed the coding principles with the other two raters. To answer questions in the reading passages, the stem of each question, the correct answer, and the required information in the reading passages to answer the question are processed in order to classify the items into a major and even sub type of ESL/EFL reading skills. The items were thus coded according to the reading skill that the three raters believed to be instrumental at arriving at the answer. In the training session, all raters reached a consensus on how to categorize the test items.

After all raters finished the coding, the data of all the raters' categorization of items were compared to check inter-rater consistency. In addition, the raters had a meeting together in order to make a final decision on items that were classified into different types of skills. If the raters had different opinions regarding the classification of any item, the raters would look at the principles for classification together again and the researcher would explain the principles to the other two raters.

The resulting agreement rates among the three raters were relatively high, with 92.47% in the SAET data sets and 91.89% in the DRET data sets. The inter-rater consistencies between two raters, i.e., the researcher versus one of the raters (as Rater 1 and Rater 2) or Rater 1 versus Rater 2, were then calculated to see if a higher consistency could be found. The inter-rater consistency rates between the researcher and Rater 2 were 98.92% in the SAET data sets and 95.95%, which were the highest agreement rates, followed by those between the researcher and Rater 1 (i.e., 92.47% in the SAET data sets and 94.59% in the DRET data sets) and those between Rater 1 and Rater 2 (i.e., 91.4% on the SAET data sets and 93.24% on the DRET data sets).

The coding results were finalized through discussion among the three raters. Since the agreement rates among the three raters were high, there is no need to exclude the coding data from any the raters and another rater. Therefore, the coding data given by all three raters were included in the data analysis. After the entire coding process was completed, seven types of reading skills were found on the 2002-2007 SAET while eight types of reading skills were found on the 2002-2007 DRET. The coding results were presented in Appendix 2.

Analysis of the Data Coded and the Passing Rates

Both qualitative and quantitative data analyses were performed to answer the research questions. To address the first and the second research questions concerning the types of reading skills measured, the qualitative analysis was conducted by adopting the revised Nuttall's taxonomy to analyze the reading comprehension items on both SAET and DRET. The frequency of reading skills will be calculated to see if there is any difference between the reading skills measured on the SAET and DRET, particularly, the present study aims to see what skills were tested the most and what skills were tested the least on both tests.

To address the third and the fourth research questions concerning the performance of all examinees and those of high achievers and low achievers, a statistical computer package (SPSS 13.0) was employed to analyze the quantitative data obtained. A two-way ANOVA with alpha level $<.05$ were conducted to see whether the differences on passing rates among test items measuring different reading skills was significant, and to see whether these differences were consistent throughout the years. In the ANOVA analysis, the reading skills measured and the year of the tests were treated as two independent variables. These two variables were examined to see if there is a pattern of reading skill occurrence throughout the years.

Summary

In this chapter, the method employed in the present study was reported. The materials, participants and instrument (i.e. the revised Nuttall's taxonomy), the qualitative (i.e., data coding procedure) and quantitative analyses (i.e., statistical analysis for the data coded) were presented.

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

This chapter presents the results and the data analysis of the present study. There are four sections in this chapter. The first section presents both qualitative and quantitative analyses of the reading skills measured on both SAET and DRET from 2002 to 2007. In the second section, the overall performances of all examinees on each item type are analyzed. A comparison of performances of the high achievers and low achievers on various item types is reported in the third section. The last section presents a summary of this chapter.

Qualitative and Quantitative Data Analysis and Results

The Reading Skills Measured in SAET & DRET

In this section, the major findings concerning the research questions presented previously are reported and discussed. A total of 167 reading comprehension test items, 93 on the SAET and 74 on the DRET, were analyzed by utilizing the revised Nuttall's taxonomy as the coding scheme. Each test item was classified according to the reading skill tested. Frequencies were run to investigate the frequency and distribution of reading level and skill type of reading comprehension test item in the SAET and the DRET. Table 6 presents the results of the items classified into two reading levels in both SAET and DRET from 2002 to 2007. According to Table 6, 70.7% aimed at testing bottom-up skills while 29.3% emphasized measuring top-down skills.

Table 6. Reading Level Tested in the SAET & DRET

Reading Level	Count	%
1. Reading for Plain Sense (Bottom-Up Skills)	118	70.7%
2. Reading into Discourse (Top-Down Skills)	49	29.3%
Total	167	100.0%

The results of item frequency on both SAET and DRET illustrated in Table 7 showed that, in general, “Recognizing and Interpreting Details” is the more frequent item on 2002 to 2007 SAET and DRET while “Recognizing Text Organization” is the least tested. In terms of readings skills at local and global level, more than half of the items of the items under analysis (around 70.7%) aimed at measuring the examinees’ reading skills at a more local level, with 9% on items testing “Word Inference from Context, 3% on items testing “Recognizing Cohesive Devices,” and 58.7% on items testing “Recognizing and Interpreting Details.” As shown in Table 6, Only 29.3% were dedicated to testing reading skills at the global level, with items on “Recognizing Functional Value,” 0.6% on “Recognizing Text Organization,” 3.6% on “Recognizing Presuppositions Underlying the Text,” 9% on “Recognizing Implications and Making Inferences,” and 8.4% on “Recognizing and Understanding the Main Idea.”

Table 7. The Frequencies of Reading Skills Measured in the SAET & the DRET

Item Type	Count	%
Skills necessary for reading for plain sense (Bottom-up Skills)		
1. Word Inference from Context	15	9.0%
2. Recognizing Cohesive Devices	5	3.0%
3. Recognizing and Interpreting Details	98	58.7%
Skills necessary for reading beyond plain sense and into discourse (Top-down skills)		
4. Recognizing Functional Value	13	7.8%
5. Recognizing Text Organization	1	0.6%
6. Recognizing Presuppositions Underlying the Text	6	3.6%
7. Recognizing Implications and Making Inferences	15	9.0%
8. Recognizing and Understanding the Main Idea	14	8.4%
Total	167	100.0%

Note. The percentages were calculated based on the total number of reading comprehension test items on 2002-2007 SAET and 2002-2007 DRET.

Frequencies were run separately to SAET and DRET to investigate the similarities and differences in the frequency and distribution of each reading level and

skill type of reading comprehension test item between the SAET and the DRET from 2002 to 2007. Table 8 presents the results regarding the reading level tested in the SAET & the DRET from 2002 to 2007.

Similarities and Differences between the SAET & the DRET

According to Table 8, in the SAET, 62.4% tested bottom-up skills while only 37.6% were devoted to measuring top-down skills. As for the DRET, 60.8% aimed at measuring bottom-up skills while only 39.2% were dedicated to top-down skills.

Reading Skills Tested in the SAET. As shown in Table 9, the research data revealed that only seven types of reading skills were identified on 2002-2007 SAET,

Table 8. Reading Level Tested in the SAET & DRET from 2002 to 2007

Category	Reading Level	Count	%
SAET	1. Reading for Plain Sense (Bottom-Up Skills)	68	73.1
	2. Reading into Discourse (Top-Down Skills)	25	26.9
	Total	93	100.0
DRET	1. Reading for Plain Sense (Bottom-Up Skills)	50	67.6
	2. Reading into Discourse (Top-Down Skills)	24	32.4
	Total	74	100.0

including “Word Inference from Context,” “Recognizing and Interpreting Details,” “Recognizing and Interpreting Details,” “Recognizing Functional Value,” “Recognizing Presuppositions Underlying the Text,” “Recognizing Implications and Making Inferences,” and “Recognizing and Understanding the Main Idea.” The most frequent item tested in the SAET is “Recognizing and Interpreting Details” (61.3%, 57 items). Items on “Word Inference from Context” came second (10.8%, 10 items), followed by items on “Recognizing and Understanding the Main Idea” (8.6%, 8 items). Items on “Recognizing Functional Value” (7.5%, 7 items) and items on “Recognizing Implications and Making Inferences” (7.5%, 7 items) were the third frequent items, followed by a very small number of occurrences of items on

“Recognizing Presuppositions Underlying the Text” (3.2%, 3 items). Only one item was coded as “Recognizing Cohesive Devices” (1.1%). On 2002 to 2007 SAET, none was labeled as “Recognizing Text Organization items.” In terms of the skills on the SAET, around 73.1% aimed at measuring skills at the local level. As for skills at the global level, only 26.9% of items tested reading skills at the global level.

Reading Skills Tested in the DRET. As illustrated in Table 9, items measuring “Recognizing and Interpreting Details” were the most frequent item type (55.4%, 41 items). Items labeled as “Recognizing Implications and Making Inferences” (10.8%, 8 items) came the second. The third frequent item type contained “Recognizing Functional Value” (8.1%, 6 items) and “Recognizing and Understanding the Main Idea” (8.1%, 6 items), followed by a very small number of occurrences of items on

Table 9. Reading Skills Measured in both SAET & DRET

Item Type	SAET		DRET	
	Count	%	Count	%
Skills necessary for reading for plain sense (Bottom-up Skills)				
1. Word Inference from Context	10	10.8%	5	6.8%
2. Recognizing Cohesive Devices	1	1.1%	4	5.4%
3. Recognizing and Interpreting Details	57	61.3%	41	55.4%
Skills necessary for reading beyond plain sense and into discourse (Top-down skills)				
4. Recognizing Functional Value	7	7.5%	6	8.1%
5. Recognizing Text Organization	0	0%	1	1.4%
6. Recognizing Presuppositions Underlying the Text	3	3.2%	3	4.1%
7. Recognizing Implications and Making Inferences	7	7.5%	8	10.8%
8. Recognizing and Understanding the Main Idea	8	8.6%	6	8.1%
Total	93	100.0%	74	100.0%

“Recognizing Cohesive Devices” (5.4%, 4 items) and items on “Recognizing Presuppositions Underlying the Text” (4.1%, 3 items). Only one item was coded as

“Recognizing Text Organization” (1.4%, 1 item), which is the least frequent item type on 2002-2007 DRET. Different from the results of the items on 2002-2007 SAET, all the eight skills on the revised Nuttall’s taxonomy were found on the DRET items from 2002 to 2007. “Recognizing Text Organization,” which was not found on 2002-2007 SAET, was measured on the DRET. Similar to the SAET, test items in the DRET also focused on measuring students’ abilities to recognize and interpret details in the passages while placing less emphasis on testing higher-order skills, which require readers to have a more global understanding of texts. Overall, 67.6% aimed at testing local reading skills while only 32.4% measured global reading skills.

Similarities and Differences between the Reading Skills Measured in the SAET & the DRET

Similarities. As shown in Table 8, in terms of the reading level tested, both examinations revealed a similar pattern: more than 60% of items emphasized on testing the examinees’ ability to use bottom-up reading skills. Much fewer percentages were used to measure the examinee’s ability to employ more top-down skills. In addition, as shown in Table 9, in terms of frequency, both examinations revealed a similar pattern: the most frequent items were items on “Recognizing and Interpreting Details” (61.3% on the SAET & 55.4% on the DRET), with much less number of items devoted to measuring other skills. Similarly, both exams had more than 60% of items testing local level reading skills. In addition, in both examinations, three types of items had occurrences of less than five times over the past six years, including items on “Recognizing Cohesive Devices” (1.1 %, 1 item in the SAET and 5.4%, 4 items in the DRET), items on “Recognizing Presuppositions Underlying the Text” (3.2%, 3 items in the SAET and 4.1%, 3 items on the DRET), and items on “Recognizing Text Organization” (0% in the SAET and 1.4%, 1 item in the DRET). This indicated that in both examinations, certain types of reading skills received less

emphasis. As can be seen from Table 8, the distribution between the two examinations showed that both examinations consisted of at least three item types each year, ranging from three to eight among the six years. In terms of the distribution of items in the SAET, there were five types in 2002, five types in 2003, three types in 2004, four types in 2005, five types in 2006, and six types in 2007. As for the distribution of items in the DRET, eight types were found in 2002, six types in 2003, four types in 2004, three types in 2005, five types in 2006, and six types in 2007. Hence, it's hard to determine whether there was a pattern of skill distribution as the skill types fluctuated year by year.

Differences. Based on close examination on the frequency of item in each year, three kinds of major differences between the SAET and the DRET were found: (1) frequency, (2) occurrences and (3) distribution of the reading skills. Firstly, in terms of frequency as shown in Table 7, the most noticeable difference between the SAET and the DRET is that the SAET had more items on “Recognizing and Interpreting Details than the DRET,” 5.9% more on the overall percentage. In addition, the SAET also had more items on “Word Inference from Context” than the DRET, 5% more on the overall percentage. To examine closely the differences between the two types of examinations, the frequency of item types year by year were further examined, which were shown in Table 10. Although the most frequent item in both exams was “Recognizing and Interpreting Details” item, more than 50% of items were devoted to measuring “Recognizing and Interpreting Details” in the SAET each year while in the DRET only five years (2002, 2003, 2004, 2006, 2007) had more than 50% of items on “Recognizing and Interpreting Details.” The frequency of “Recognizing and Interpreting Details” item dropped slightly to 45.5% in 2005 DRET. Regarding occurrences of reading skills shown in Table 8, “Recognizing Text Organization” was

tested in the DRET (2002) but not tested in the SAET. Items which appeared every year in 2002-2007 SAET are those on “Word Inference from Context” and “Recognizing and Interpreting Details” whereas only those on “Recognizing and Interpreting Details” appeared consistently in the DRET from 2002 to 2007. In addition, items on “Recognizing Implications and Making Inferences” occurred nearly every year except in 2004 SAET and in 2005 DRET. Moreover, items on “Recognizing and Understanding the Main Idea” occurred nearly every year except in the 2005 SAET; while item of this type only occurred in the 2002, 2003, and 2005 DRET. Thus, the results indicated that certain items were favored in the SAET: e.g., items on “Word Inference from Context,” “Recognizing and Interpreting Details,” “Recognizing Implications and Making Inferences” and “Recognizing and Understanding the Main Idea.” On the other hand, the DRET had more items on “Recognizing and Interpreting Details” (occurred every year), “Word Inference from Context” (occurred in 2002, 2003, 2005, 2006) and “Recognizing Functional Value” (occurred in 2002, 2003, 2004, and 2006). Another difference between the two examinations was in the distribution of item types for each year. The results showed that on the SAET, two item types occurred each year, including items on “Word Inference from Context,” and items on “Recognizing and Interpreting Details.” As to the DRET, only items on “Recognizing and Interpreting Details” were tested every year. In sum, the results indicated items on “Recognizing and Interpreting Details” were the majority on both SAET and DRET. Thus, it indicated that both exams emphasized measuring examinees’ abilities to identify and interpret specific information in the reading passages, which belonged to the more local level comprehension.

Table 10. Reading Skills Measured in 2002-2007 SAET & DRET

SAET												
Item Type	2002		2003		2004		2005		2006		2007	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Skills necessary for Reading for Plain Sense (Bottom-up Skills)												
1. Word Inference from Context	2	13.3%	1	6.7%	2	13.3%	1	6.3%	2	12.5%	2	12.5%
2. Recognizing Cohesive Devices	1	6.7%	0	.0%	0	.0%	0	.0%	0	.0%	0	.0%
3. Recognizing and Interpreting Details	9	60.0%	8	53.3%	11	73.3%	10	62.5%	10	62.5%	9	56.3%
Skills necessary for Reading Beyond the Plain and into Discourse (Top-down Skills)												
4. Recognizing Functional Value	0	.0%	0	.0%	0	.0%	4	25.0%	2	12.5%	1	6.3%
5. Recognizing Text Organization	0	.0%	0	.0%	0	.0%	0	.0%	0	.0%	0	.0%
6. Recognizing Presuppositions Underlying the Text	0	.0%	1	6.7%	0	.0%	0	.0%	0	.0%	2	12.5%
7. Recognizing Implications and Making Inferences	2	13.3%	2	13.3%	0	.0%	1	6.3%	1	6.3%	1	6.3%
8. Recognizing and Understanding the Main Idea	1	6.7%	3	20.0%	2	13.3%	0	.0%	1	6.3%	1	6.3%
Total	15	100.0%	15	100.0%	15	100.0%	16	100.0%	16	100.0%	16	100.0%
DRET												
Item Type	2002		2003		2004		2005		2006		2007	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Skills necessary for Reading for Plain Sense (Bottom-up Skills)												
1. Word Inference from Context	1	6.7%	1	6.7%	1	9.1%	0	.0%	0	.0%	0	.0%
2. Recognizing Cohesive Devices	1	6.7%	2	13.3%	1	9.1%	2	18.2%	1	9.1%	0	.0%
3. Recognizing and Interpreting Details	8	53.3%	7	46.7%	6	54.5%	5	45.5%	6	54.5%	8	72.7%
Skills necessary for Reading Beyond the Plain and into Discourse (Top-down Skills)												
4. Recognizing Functional Value	1	6.7%	2	13.3%	2	18.2%	0	.0%	1	9.1%	0	.0%
5. Recognizing Text Organization	1	6.7%	0	.0%	0	.0%	0	.0%	0	.0%	1	9.1%
6. Recognizing Presuppositions Underlying the Text	1	6.7%	0	.0%	0	.0%	0	.0%	1	9.1%	2	18.2%
7. Recognizing Implications and Making Inferences	1	6.7%	2	13.3%	1	9.1%	0	.0%	2	18.2%	0	.0%
8. Recognizing and Understanding the Main Idea	1	6.7%	1	6.7%	0	.0%	4	36.4%	0	.0%	0	.0%
Total	15	100.0%	15	100.0%	11	100.0%	11	100.0%	11	100.0%	11	100.0%

Examinees' Performance on Each Question Type

As shown in the preceding section, eight types of skills were measured in 2002-2007 SAET and DRET. To analyze the examinees' overall performance on each reading skill in both SAET and DRET, the passing rates on each item type were examined. A two-way ANOVA was applied to the SAET and DRET to see if there were significant differences on students' average passing rates among various question types, and to see whether these differences were consistent throughout the six years. The passing rates are used to show the proportion of the examinees who got an item correct. In the ANOVA analyses, the item types and the year of exams were used as the factors. Descriptive statistics for the passing rates on each item type on the SAET from 2002 to 2007 and the results of ANOVA are presented in the following section, followed by those of the DRET.

SAET

Descriptive statistics for the passing rates on each reading comprehension test item from 2002 to 2007 SAET are summarized in Table 9. Table 11 shows that, in the 2002 SAET, only five types of reading skills were identified and the total mean of the passing rate was 60.8 (SD=12.33). The passing rates for "Recognizing and Understanding the Main Idea" items were the highest of all, which belonged to global reading skill types, followed by "Recognizing Cohesive Devices" items (Mean=75), "Recognizing and Interpreting Details" items (Mean=61.1111, SD=10.9), "Word Inference from Context" items (Mean=58, SD=1.41), and "Recognizing Implications and Making Inferences" items (Mean=45, SD=7.07).

In the 2003 SAET, five types of reading skills were identified. The total mean of the passing rates was 51.3333. Items on "Recognizing and Understanding the Main Idea," a global reading skill, received the highest passing rates (Mean=61.3333),

Table 11. Average Passing Rates of Items Measuring Different Reading Skills in 2002-2007 SAET and DRET

SAET												
Item Type	2002		2003		2004		2005		2006		2007	
	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
Skills necessary for Reading for Plain Sense (Bottom-up Skills)												
1. Word Inference from Context	58.0000	2	44.0000	1	61.0000	2	58.0000	1	41.0000	2	43.0000	2
2. Recognizing Cohesive Devices	75.0000	1	N/A	0	N/A	0	N/A	0	N/A	0	N/A	0
3. Recognizing and Interpreting Details	61.1111	9	50.7500	8	58.6364	11	55.2000	10	59.7000	10	60.4444	9
Skills necessary for Reading Beyond the Plain and into Discourse (Top-down Skills)												
4. Recognizing Functional Value	N/A	0	N/A	0	N/A	0	54.5000	4	40.5000	2	67.0000	1
5. Recognizing Text Organization	N/A	0	N/A	0	N/A	0	N/A	0	N/A	0	N/A	0
6. Recognizing Presuppositions Underlying the Text	N/A	0	51.0000	1	N/A	0	N/A	0	N/A	0	54.5000	2
7. Recognizing Implications and Making Inferences	45.0000	2	41.0000	2	N/A	0	66.0000	1	76.0000	1	41.0000	1
8. Recognizing and Understanding the Main Idea	81.0000	1	61.3333	3	54.0000	2	N/A	0	N/A	0	77.0000	1
Total	60.8000	15	51.3333	15	58.3333	15	55.8750	16	56.0000	16	57.7500	16
DRET												
Item Type	2002		2003		2004		2005		2006		2007	
	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
Skills necessary for Reading for Plain Sense (Bottom-up Skills)												
1. Word Inference from Context	54.0000	1	48.0000	1	N/A	0	50.5000	2	31.0000	1	N/A	0
2. Recognizing Cohesive Device	57.0000	1	43.0000	1	44.5000	2	N/A	0	N/A	0	N/A	0
3. Recognizing and Interpreting Details	45.5000	8	46.3750	8	37.8333	6	46.4000	5	39.0000	6	49.3750	8
Skills necessary for Reading Beyond the Plain and into Discourse (Top-down Skills)												
4. Recognizing Functional Value	51.0000	1	54.5000	2	48.0000	2	N/A	0	68.0000	1	N/A	0
5. Recognizing Text Organization	50.0000	1	N/A	0	N/A	0	N/A	0	N/A	0	N/A	0
6. Recognizing Presuppositions Underlying the Text	30.0000	1	N/A	0	N/A	0	N/A	0	28.0000	1	29.0000	1
7. Recognizing Implications and Making Inferences	31.0000	1	30.5000	2	27.0000	1	N/A	0	45.5000	2	40.0000	2
8. Recognizing and Understanding the Main Idea	42.0000	1	49.0000	1	N/A	0	56.7500	4	N/A	0	N/A	0
Total	45.2667	15	45.4000	15	39.9091	11	40.1000	11	41.0909	11	45.8182	11

followed by items on “Recognizing Presuppositions Underlying the Text” (Mean=51), items on “Recognizing and Interpreting Details” (Mean=50.75), items on “Word Inference from Context” (Mean=44), and items on “Recognizing Implications and Making Inferences” (Mean=41).

As shown in Table 11, only three types of reading skills were identified in the 2004 SAET. The total mean of the passing rates was 58.33. Items on “Word Inference from Context,” which belonged to the local reading skill, were the best performed, with the highest passing rates of 61, followed by items on “Recognizing and Interpreting Details” (Mean=58.64), and items on “Recognizing and Understanding the Main Idea“(Mean=54).

In the 2005 SAET, only four types of reading skills were identified. The total mean of the passing rates was 55.88. “Recognizing Implications and Making Inferences” items, which belonged to the global reading skill level, received the highest passing rates (Mean=66), followed by “Word Inference from Context” items (Mean=58), “Recognizing and Interpreting Details” items (Mean=55.2), and “Recognizing Functional Value” (Mean=54.5).

Table 11 also shows that, in 2006 SAET, only four types of reading skills were identified. The total mean of the passing rate was 56. “Recognizing Implications and Making Inferences” items, which was a global reading skill, had the highest passing rates (Mean=76). This was followed by “Recognizing and Interpreting Details” items (Mean=59.7), “Word Inference from Context” (Mean=41), and “Recognizing Functional Value “(Mean=40.5).

In the 2007 SAET, 6 types of reading skills were identified, and the total mean of the passing rate was 57.75. Items on “Recognizing and Understanding the Main Idea”, which was a global reading skill, received the highest passing rates (Mean=77),

followed by items on “Recognizing Functional Value” (Mean=67), items on “Recognizing and Interpreting Details” (Mean=60.44), items on “Recognizing Presuppositions Underlying the Text” (Mean=54.5), items on “Word Inference from Context” (Mean=43), and items on “Recognizing Implications and Making Inferences” (Mean=41).

In general, in the SAET from 2002 to 2007, the ranking of passing rates among all item types were not consistent each year. In 2002, 2003 and 2007, items on “Recognizing and Understanding the Main Idea” ranked the highest, while items on “Word Inference from Context” ranked the highest in 2004, and items on “Recognizing Implications and Making Inferences” ranked the highest in 2005 and 2006. In other words, from 2002 to 2007, more global-level reading skills tend to have the highest passing rates. As for the worst performed items each year, items on “Recognizing Implications and Making Inferences” ranked the lowest in and 2002, 2003 and 2007 while items on “Recognizing and Understanding the Main Idea” ranked the lowest in 2004, and items on “Recognizing Functional Value” ranked the lowest in 2005 and 2006. The results showed that from 2002 to 2007, the items which had the lowest passing rates all tested global reading skill. One thing worth noting is that items on “Recognizing and Understanding the Main Idea” had the highest passing rates in 2002, 2003, and 2007 while ranked the lowest in 2004.

To yield statistical support to these observations about the inconsistent passing rates given by each item type throughout 2002-2007, a two-way ANOVA was conducted. This is done to see whether the differences on passing rates among items measuring different reading skills were significant, and to see whether these differences exhibited consistency throughout the years. The results of the ANOVA analysis were show in Table 12. The results revealed that there was no significant

effect on the variable of the item type ($F=1.169, p>.05$), which indicated the passing rates did not vary according to the types of skills measured. In addition, there was no effect on the variable of the year ($F=.728, p>.05$), which indicated that the passing rates do not vary throughout the years. Also, no significant effect was found on the variable of the interaction of item type and year ($F=.927, p>.05$), which indicated that the discrepancy of the average passing rates on each item type would not differ throughout years and each of these years showed a similar pattern. Furthermore, from Table 11, it can also be seen that different item types appeared in different years. Hence, it's hard to find a general pattern of item types from 2002 to 2007 SAET.

Table 12. The ANOVA Analysis of the Passing Rates in the SAET

Dependent Variable: Passing Rates

Category	Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SAET	Corrected Model	5401.767(a)	27	200.065	.991	.494
	Intercept	113352.846	1	113352.846	561.259	.000
	Item Type	1416.596	6	236.099	1.169	.334
	Year	735.227	5	147.045	.728	.605
	Item Type * Year	2996.387	16	187.274	.927	.543
	Error	13127.523	65	201.962		
	Total	316936.000	93			
	Corrected Total	18529.290	92			

a. R Squared = .292 (Adjusted R Squared = -.003)

DRET

The descriptive analysis on the average passing rates on each item type of the DRET from 2002 to 2007 is presented in Table 11. As illustrated in Table 11, in 2002 DRET, the total mean of the passing rates was 45.27. “Recognizing Cohesive Devices” items received the highest mean (Mean=57), followed by “Word Inference from Context” items (Mean=54), “Recognizing Functional Value” items (Mean=51), “Recognizing Text Organization” items (Mean=50), “Recognizing and Interpreting Details” items (Mean=45.5), “Recognizing and Understanding the Main Idea” (Mean=42), “Recognizing Implications and Making Inferences” (Mean=31), and

“Recognizing Presuppositions Underlying the Text” items had the lowest passing rates (Mean=30).

For the 2003 DRET (as shown in Table 11), the total mean of the passing rates was 45.40. Items on “Recognizing Functional Value” were the best-performed, with a mean of 54.5 (SD=26.16), followed by items on “Recognizing and Understanding the Main Idea” (Mean=49), items on “Word Inference from Context” (Mean=48), items on “Recognizing and Interpreting Details” (Mean=46.38), items on “Recognizing Cohesive Devices” (Mean=43), and items on “Recognizing Implications and Making Inferences” (Mean=30.5), which got the lowest average passing rates.

As shown in Table 11, in 2004 DRET, the total mean of the average passing rates was 39.9091. Items on “Recognizing Functional Value” (Mean=48) obtained the highest passing rates, followed by items on “Recognizing Cohesive Devices” (Mean=44.5), items on “Recognizing and Interpreting Details” (Mean=37.83). Items on “Recognizing Implications and Making Inferences” had a quite low passing rate of 27 only, the lowest in 2004 DRET and also the lowest in the DRETs over the past 6 years.

From Table 11, it can be seen that, in the 2005 DRET, the total mean of the average passing rates was 40.1. Items on “Recognizing and Understanding the Main Idea” (Mean=56.75) received the highest passing rate, followed by items on “Word Inference from Context” (Mean=50.5), and items on “Recognizing and Interpreting Details” (Mean=46.4).

As shown in Table 11, in 2006 DRET, the total mean of the average passing rates was 41.09. Items on “Recognizing Functional Value” (Mean=68) got the highest passing rate, followed items on “Recognizing Implications and Making Inferences” (Mean 45.5). Items on “Recognizing and Interpreting Details” (Mean=39) ranked the

third, followed by items on Word Inference from *Context* (Mean=31). Items on “Recognizing Presuppositions Underlying the Text” got the lowest passing rate of 28.

As illustrated in Table 11, in 2007 DRET, the total mean of the average passing rates was 45.82. Items on “Recognizing and Interpreting Details” (Mean=39) were the best performed, followed by items on “Recognizing Implications and Making Inferences” (Mean=40) and items on “Recognizing Presuppositions Underlying the Text” were the worst performed.

In sum, in the DRET from 2002 to 2007, similar to the results in the SAET, the ranking of passing rates of different item types were not consistent throughout the years. Items on “Recognizing Cohesive Devices” ranked the highest in 2002, items on “Recognizing and Understanding the Main Idea” ranked the highest in 2003 and 2005, while items on “Recognizing Functional Value” ranked the highest in 2004 and 2006 and items on “Recognizing and Interpreting Details” ranked the highest in 2007. As for items that ranked lastly, items on “Recognizing Presuppositions Underlying the Text” ranked the lowest in 2002, 2006, and 2007, items on “Recognizing Implications and Making Inferences” ranked the lowest in 2003 and 2004, and items on “Recognizing and Interpreting Details” ranked lastly in 2005. Thus, it’s difficult to determine whether the reading skills tested in the DRET exhibited certain patterns.

To yield statistical support to these observations about the inconsistent passing rates given by each item type throughout 2002-2007, a two-way ANOVA was conducted. This is done to see whether the differences on passing rates among items measuring different reading skills were significant, and to see whether these differences exhibited consistency throughout the years. The results of the ANOVA analysis of the DRET were shown in Table 13.

Table 13. The ANOVA Analysis of Passing Rates in the DRET

Dependent Variable: Passing Rates

Category	Source	Type III Sum of Squares	df	Mean Square	F	Sig.
DRET	Corrected Model	4394.926(a)	28	156.962	.957	.540
	Intercept	66411.932	1	66411.932	405.114	.000
	Item Type	2200.833	7	314.405	1.918	.089
	Year	513.169	5	102.634	.626	.681
	Item Type * Year	1324.555	16	82.785	.505	.931
	Error	7377.033	45	163.934		
	Total	160275.000	74			
	Corrected Total	11771.959	73			

a R Squared = .373 (Adjusted R Squared = -.017)

The results of two-way ANOVA for the DRET were similar as those of the SAET. The results of the ANOVA showed that there was no effect on the variable of item type ($F=1.918, p>.05$), the year ($F=.626, p>.05$), and the interaction between the item type and year ($F=.505, p>.05$). Thus, the ANOVA analyses also indicated that in the DRET, the passing rates did not vary according to the types of reading skills measured on each time, the year, or the interaction of item types and year. However, if we examine the passing rates of different item types throughout 2002 to 2007 DRET, the passing rates of item types were not consistent throughout the years. For example, “Word Inference from Context” had a passing rate of 54 in 2002, 48 in 2003, 50.5 in 2005, but dropped to only 31 in 2006 DRET. In addition, not all types of skills were the most or the least difficult ones. Since different kinds of skills appeared in different years, the passing rates of the specific skills cannot be compared.

Comparisons of High & Low Achievers’ Performances on Different Item Types

In this section, the passing rates of the high achievers (Ph) and low achievers (Pl), and the discrimination indexes (i.e., $D=Ph-Pl$) on each reading level and item type were examined to explore the differences between the high and low achievers on

performances of different test items. Since the discrimination index (D) represents the discrepancy between the passing rates of the high achievers and low achievers, one-way ANOVA was run to examine how different reading levels and item types discriminated the high achievers from the low achievers. The following sections first report the descriptive statistics for the high and low achievers' passing rates and the discrimination index on each reading level on the SAET and the DRET and the results of one-way ANOVA both tests, and then proceed to those of each item type on both exams. Table 12 showed the passing rates of the high achievers and low achievers and the discrimination index in both exams.

Results of the SAET

As illustrated in Table 15, in terms of the two reading levels in the SAET, high achievers performed quite well on both bottom-up skills and top-down skills, with passing rates of more than 80 for both levels. Low achievers performed poorly on either bottom-up or top-down skills, with passing rates of around 30 for each level. One thing worth noting is that the passing rates of low achievers on top-down skills were slightly below the standard set by Jeng et al. (1999). Jeng et al. (1999) suggested that the idea item facility index for national examination such as college entrance examination should be 30 to 80. To look at the performance of high and low achievers on different types of reading skills, the passing rates of high and low achievers on each type of skill were examined. The results the passing rates of high and low achievers and the discrimination indexes on different reading levels in the SAET and DRET were presented in the following table.

Table 14. Passing Rates of High & Low Achievers and the Discrimination Indexes on Different Reading Levels in the SAET & DRET

Category			N	Mean	SD
SAET	Ph	1. Reading for Plain Sense (Bottom-Up Skills)	58	85.0172	13.08205
		2. Reading into Discourse (Top-Down Skills)	35	80.1429	12.74672
		Total	93	83.1828	13.10405
	Pl	1. Reading for Plain Sense (Bottom-Up Skills)	58	31.2586	12.74247
		2. Reading into Discourse (Top-Down Skills)	35	29.6286	10.28158
		Total	93	30.6452	11.84469
	D	1. Reading for Plain Sense (Bottom-Up Skills)	58	53.7586	11.10604
		2. Reading into Discourse (Top-Down Skills)	35	50.5143	10.70561
		Total	93	52.5376	11.01240
DRET	Ph	1. Reading for Plain Sense (Bottom-Up Skills)	45	72.2000	15.71566
		2. Reading into Discourse (Top-Down Skills)	29	68.5172	18.48478
		Total	74	70.7568	16.82856
	Pl	1. Reading for Plain Sense (Bottom-Up Skills)	45	22.2222	7.47082
		2. Reading into Discourse (Top-Down Skills)	29	24.1724	8.76303
		Total	74	22.9865	8.00084
	D	1. Reading for Plain Sense (Bottom-Up Skills)	45	49.9778	13.60396
		2. Reading into Discourse (Top-Down Skills)	29	44.3448	15.55698
		Total	74	47.7703	14.56168

Note. Ph refers to passing rates generated by high achievers.

Pl refers to passing rates generated by low achievers.

D refers to the discrimination index (i.e., Ph-Pl).

The high and low achievers refer to the examinees who reach the top and the bottom 33% in the population of the SAET and DRET.

Table 15. Passing Rates of the High & Low Achievers and the Discrimination Indexes on Different Item Types in 2002-2007 SAET

Year	Item Type	Ph		Pl		D		N
		Mean	SD	Mean	SD	Mean	SD	
2002	1. Word Inference from Context	88.0000	.00000	28.5000	4.94975	59.5000	4.94975	2
	2. Recognizing Cohesive Devices	99.0000	.	33.0000	.	66.0000	.	1
	3. Recognizing and Interpreting Details	89.0000	7.44983	32.6667	11.96871	56.3333	13.53699	9
	7. Recognizing Implications and Making Inferences	64.0000	2.82843	25.0000	8.48528	39.0000	5.65685	2
	8. Recognizing and Understanding the Main Idea	97.0000	.	54.0000	.	43.0000	.	1
	Total	86.7333	11.31034	32.5333	11.49451	54.2000	12.94604	15
2003	1. Word Inference from Context	76.0000	.	17.0000	.	59.0000	.	1
	3. Recognizing and Interpreting Details	79.7500	14.25031	24.8750	10.34322	54.8750	10.50765	8
	6. Recognizing Presuppositions Underlying the Text	87.0000	.	18.0000	.	69.0000	.	1
	7. Recognizing Implications and Making Inferences	59.5000	6.36396	27.5000	.70711	32.0000	7.07107	2
	8. Recognizing and Understanding the Main Idea	88.0000	11.13553	31.3333	8.73689	56.6667	4.50925	3
	Total	78.9333	14.11922	25.5333	9.03854	53.4000	12.25794	15
2004	1. Word Inference from Context	91.5000	6.36396	27.5000	.70711	64.0000	7.07107	2
	3. Recognizing and Interpreting Details	83.2727	14.02206	33.3636	13.42589	49.9091	12.01212	11
	8. Recognizing and Understanding the Main Idea	75.0000	14.14214	37.5000	4.94975	37.5000	9.19239	2
	Total	83.2667	13.30664	33.1333	11.74045	50.1333	12.76640	15

2005	1. Word Inference from Context	85.0000	.	27.0000	.	58.0000	.	1
	3. Recognizing and Interpreting Details	80.0000	17.24980	31.1000	12.70564	48.9000	10.54567	10
	4. Recognizing Functional Value	81.5000	9.18332	28.5000	7.76745	53.0000	6.83130	4
	7. Recognizing Implications and Making Inferences	91.0000	.	37.0000	.	54.0000	.	1
	Total	81.3750	14.27293	30.5625	10.67064	50.8125	9.15947	16
2006	1. Word Inference from Context	62.0000	14.14214	24.0000	9.89949	38.0000	4.24264	2
	3. Recognizing and Interpreting Details	88.0000	10.54093	31.8000	16.73187	56.2000	11.38029	10
	4. Recognizing Functional Value	71.5000	17.67767	19.5000	10.60660	52.0000	7.07107	2
	7. Recognizing Implications and Making Inferences	97.0000	.	44.0000	.	53.0000	.	1
	8. Recognizing and Understanding the Main Idea	91.0000	.	32.0000	.	59.0000	.	1
	Total	83.4375	14.52799	30.0625	14.73078	53.3750	11.02044	16
2007	1. Word Inference from Context	71.5000	7.77817	23.0000	1.41421	48.5000	6.36396	2
	3. Recognizing and Interpreting Details	88.5556	12.23837	32.3333	12.12436	56.2222	7.49630	9
	4. Recognizing Functional Value	83.0000	.	45.0000	.	38.0000	.	1
	6. Recognizing Presuppositions Underlying the Text	84.5000	6.36396	31.0000	11.31371	53.5000	4.94975	2
	7. Recognizing Implications and Making Inferences	78.0000	.	14.0000	.	64.0000	.	1
	8. Recognizing and Understanding the Main Idea	95.0000	.	54.0000	.	41.0000	.	1
	Total	85.3125	11.34148	32.0000	12.77498	53.3125	8.70799	16

Note. Table 13 continued.

Ph refers to passing rates generated by high achievers.

Pl refers to passing rates generated by low achievers.

D refers to the discrimination index. (i.e., Ph-Pl).

As shown in Table 15, in the 2002 SAET, the high achievers performed well on all types of item types while low achievers performed quite well on “Recognizing and Understanding the Main Idea” items only. The high achievers performed extremely well on items measuring “Recognizing Cohesive Devices” items (Mean=99), followed closely by items on “Recognizing and Understanding the Main Idea” (Mean=97), while the low achievers performed best on items measuring “Recognizing and Understanding the Main Idea” (Mean=54). As to items that were worst performed, both groups performed worst on items measuring “Recognizing Implications and Making Inferences” (Mean=64 vs. Mean=25 respectively for high and low achievers).

In addition, items of this type also produced the worst discriminatory power in 2002 SAET ($D=39$). For the discrepancy of these two groups, items on “Recognizing Cohesive Devices” ($D=66$) had the highest discrimination indexes. Thus items of this type best discriminated higher achiever from low achievers. As illustrated in Table 15, in the 2003 SAET, both high achievers and low achievers performed best on items measuring “Presuppositions Underlying the Text” (Mean=88 vs. Mean=31.33 respectively for high and low achievers). The high achievers performed worst on items measuring “Recognizing Implications and Making Inferences” (Mean=59) whereas the low achievers performed worst on items measuring “Word Inference from Context (Mean=19), which is far lower below the standard index set by Jeng et al (1999). In terms of the discrepancy of discrimination index, items on “Presuppositions Underlying the Text” discriminated the examinees the best ($D=69$). Similar to the results in 2002 SAET, items on “Recognizing Implications and Making Inferences” ($D=32$) had the weakest discriminatory power.

As shown in Table 15, in the 2004 SAET, it can be seen that the high achievers

performed best on items measuring “Word Inference from Context” (Mean=91.5) while the low achievers performed best on items measuring “Recognizing and Understanding the Main Idea” (Mean=37). As to the worst performed items between these two groups, the high achievers performed worst on items measuring “Recognizing and Understanding the Main Idea” (Mean=75) whereas the low achievers performed worst on items measuring “Word Inference from Context” (Mean=27.5), which is slightly below the ideal index set by Jeng et al. (1999). In 2004 SAET, items on “Word Inference from Context” (D=64) had the highest discrimination index in 2004 while items on “Recognizing and Understanding the Main Idea” had the lowest discrimination index.

As to the 2005 SAET, both high and low achievers performed best on items measuring “Recognizing Implications and Making Inferences” (Mean=91 and Mean=37 respectively). The items worst performed by the high achievers were items on “Recognizing and Interpreting Details” (Mean=80) whereas the items worst performed by the low achievers were items on “Word Inference from Context” (Mean=27), which was below the reasonable discrimination index of 30⁵ suggested by Jeng et al (1999). For the discrepancy between the two groups of examinees, the examinees, the high achievers and low achievers performed most differently on items measuring “Word Inference from Context” (D=58). That is to say, global reading skills best distinguished high and low achievers in 2005 SAET.

⁵ Jeng (1999) suggested that the minimum discrimination index for an English multiple-choice item with four options should be .30, which is slightly higher than Kuo’s (1989:271) .25 in order to have a higher standard of items.

As shown in Table 15, in 2006 SAET, both high and low achievers were good at items on “Recognizing Implications and Making Inferences” (Mean=97 vs. Mean=44 for the high and low achievers respectively). The high achievers performed worst on items measuring “Word Inference from Context” (Mean=62) while the low achievers performed worst on two types of item that had passing rates far below the reasonable discrimination index suggested by Jen et al. (1999)—“Word Inference from Context” (Mean=24) and “Recognizing Functional Value” (Mean=19.5). Both high achievers and low achievers performed poorly on global skills. The two groups of examinees performed most differently on items measuring “Recognizing and Understanding the Main Idea,” with a discrimination index of 59.

Lastly, in 2007 SAET, both high achievers and low achievers performed the best on items measuring “Recognizing and Understanding the Main Idea” (Mean=95 vs. 54 for the high and low achievers respectively). Similar to the results in 2002 SAET, low achievers had relatively high passing rates on this item, with a mean of 54. The high achievers performed the worst on items measuring “Word Inference from Context” (Mean=71.5) while the low achievers performed the worst on items measuring “Recognizing Implications and Making Inferences” (Mean=14). Items on “Recognizing and Making Inferences” (Mean=64) had the best discriminatory power. Again, global reading skills best discriminated high and low achievers in the SAET.

In general, in the SAET from 2002 to 2007, the discrimination indexes of most item types each year were far above the ideal discrimination index of 30 established by Jeng et al. (1999), with the average of discrimination index around 50. To find out whether the differences on the discrimination indexes among item types measuring different reading skills each year were significant, a two-way ANOVA test was run to analyze the data. The fixed factors are the year and item types while the dependent

variable is the discrimination indexes. As shown in Table 15, the results of the ANOVA revealed that no significant effect was found either in the item types on the discrimination index alone or on the factor of year alone. However, a significant effect was found on the interaction of item types and year on the discrimination indexes ($F=1.842, p<.05$), which suggested that there were significant differences among the discrimination indexes of different item types. That is to say, the discrimination indexes varied according to different types of reading skills measured. However, not all types of reading skills were tested each year and thus it was difficult to determine which type of reading skill best discriminated high achievers from low achievers.

Table 16. The ANOVA Analysis of Discrimination Indexes in the SAET
Dependent Variable: D

Category	Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SAET	Corrected Model	4304.612(a)	27	159.430	1.512	.089
	Intercept	96481.318	1	96481.318	915.181	.000
	Item Type	956.211	6	159.369	1.512	.188
	Year	137.192	5	27.438	.260	.933
	Item Type * Year	3106.440	16	194.152	1.842	.044
	Error	6852.506	65	105.423		
	Total	267856.000	93			
	Corrected Total	11157.118	92			
	Total	184347.000	74			
	Corrected Total	15479.095	73			

a R Squared = .386 (Adjusted R Squared = .131)

Results of the DRET

As shown in Table 17, in the 2002 DRET, the high achievers in general performed well on all item types, except items on “Recognizing Implications and Making Inferences” since the passing rate for the high achievers was only 45. On the other hand, the low achievers performed poorly on all items and passing rates of all item types were lower than the minimum desirable discrimination index of .30

suggested by Jeng et al (1999), which means that the low achievers performed poorly on both local and global skills in 2002 DRET. Items that the high achievers performed the best are items on “Recognizing Text Organization” (Mean=86) while the low achievers performed the worst on this type of question, with a mean of only 14. Items of this type also has the highest discrimination indexes (D=72). That is to say, global skill “Recognizing Text Organization” best discriminated the high and low achievers.

From the results of 2003 DRET, it can be seen that high achievers in general performed well on all items since the passing rates of all items were more than 50 and most of them reached above 74. The high achievers performed best on items measuring “Recognizing Cohesive Devices” (Mean=83) whereas the low achievers performed best on items measuring “Recognizing Functional Value” (Mean=31.5). Both groups of examinees performed worst on items measuring “Recognizing Implications and Making Inferences” (Mean=57 vs. Mean=14 for high and low achievers respectively). That is to say, items of this type were the most difficult for both high and low achievers. High and low achievers performed most differently on local reading skill “Recognizing Cohesive Devices” (D=64) in 2003.

As illustrated in 2004 DRET, both high achievers performed best on items measuring “Recognizing Functional Value” (Mean=73.5 vs. Mean=25 for high and low achievers respectively). Besides, the second best performed items for both groups of learners are items on “Recognizing Cohesive Devices” (Mean =73 vs. Mean=22). These two types of questions, one is a local skill and the other is a global skill) best discriminated the examinees, with a discriminatory power of 48.5 for each skill.

For the 2005 DRET, the high achievers obtained high passing rates on all items and they performed best on items measuring “Recognizing and Understanding the Main Idea” (Mean=84.25). The low achievers performed best on items measuring

Table 17. Passing Rates of the High & Low Achievers and the Discrimination Indexes on Different Item Types in 2002-2007 DRET

Year	Item Type	Ph		Pl		D		N
		Mean	SD	Mean	SD	Mean	SD	
2002	1. Word Inference from Context	79.0000	.	29.0000	.	50.0000	.	1
	2. Recognizing Cohesive Devices	85.0000	.	29.0000	.	56.0000	.	1
	3. Recognizing and Interpreting Details	72.3750	12.02304	23.8750	8.30555	48.5000	8.88015	8
	4. Recognizing Functional Value	85.0000	.	20.0000	.	65.0000	.	1
	5. Recognizing Text Organization	86.0000	.	14.0000	.	72.0000	.	1
	6. Recognizing Presuppositions Underlying the Text	50.0000	.	16.0000	.	34.0000	.	1
	7. Recognizing Implications and Making Inferences	45.0000	.	23.0000	.	22.0000	.	1
	8. Recognizing and Understanding the Main Idea	66.0000	.	18.0000	.	48.0000	.	1
	Total	71.6667	14.25115	22.6667	7.18795	49.0000	12.93942	15
2003	1. Word Inference from Context	82.0000	.	23.0000	.	59.0000	.	1
	2. Recognizing Cohesive Devices	83.0000	.	19.0000	.	64.0000	.	1
	3. Recognizing and Interpreting Details	76.3750	18.14180	23.3750	10.35012	53.0000	12.80625	8
	4. Recognizing Functional Value	77.0000	26.87006	31.5000	13.43503	45.5000	13.43503	2
	7. Recognizing Implications and Making Inferences	57.0000	1.41421	14.0000	5.65685	43.0000	4.24264	2
	8. Recognizing and Understanding the Main Idea	74.0000	.	22.0000	.	52.0000	.	1
		Total	74.5333	16.49618	22.6667	9.58570	51.7333	11.31665

2004	2. Recognizing Cohesive Devices	73.0000	3.53553	22.0000	11.31371	48.5000	7.77817	2
	3. Recognizing and Interpreting Details	61.1667	10.34247	20.6667	3.44480	40.5000	8.59651	6
	4. Recognizing Functional Value	73.5000	20.50610	25.0000	16.97056	48.5000	3.53553	2
	7. Recognizing Implications and Making Inferences	42.0000	.	20.0000	.	22.0000	.	1
Total		63.3636	13.26855	21.6364	7.11720	41.7273	10.12019	11
2005	1. Word Inference from Context	75.0000	16.97056	30.0000	2.82843	45.0000	14.14214	2
	3. Recognizing and Interpreting Details	77.2000	14.41180	19.6000	5.72713	57.6000	14.65606	5
	8. Recognizing and Understanding the Main Idea	84.2500	8.50000	27.7500	4.92443	56.5000	10.50397	4
Total		79.3636	12.21698	24.4545	6.59339	54.9091	12.77853	11
2006	1. Word Inference from Context	59.0000	.	17.0000	.	42.0000	.	1
	3. Recognizing and Interpreting Details	68.6667	22.86190	17.0000	4.77493	51.6667	19.99667	6
	4. Recognizing Functional Value	86.0000	.	48.0000	.	38.0000	.	1
	6. Recognizing Presuppositions Underlying the Text	35.0000	.	23.0000	.	12.0000	.	1
	7. Recognizing Implications and Making Inferences	62.5000	38.89087	30.0000	8.48528	32.5000	30.40559	2
Total		65.1818	23.65510	22.7273	10.77117	42.4545	21.39796	11
2007	3. Recognizing and Interpreting Details	73.1250	17.71551	25.7500	7.45941	47.3750	16.58689	8
	6. Recognizing Presuppositions Underlying the Text	45.0000	.	17.0000	.	28.0000	.	1
	7. Recognizing Implications and Making Inferences	63.0000	12.72792	19.5000	2.12132	43.5000	14.84924	2
Total		68.7273	17.72620	23.8182	7.12486	44.9091	15.76359	11

Note. Table 14 continued.

Ph refers to passing rates generated by the high achievers.

Pl refers to passing rates generated by the low achievers.

D refers to the discrimination index (i.e., Ph-Pl).

“Word Inference from Context” (Mean=30). However, local items on “Recognizing and Interpreting Details” (D=57.6) discriminated the examinees best, followed by global items on “Recognizing and Understanding the Main Idea” (D=56.5), and global items on “Word Inference from Context” (D=45).

As revealed in 2006 DRET, the high achievers performed well on all items except items on “Recognizing Presuppositions Underlying the Text,” with an average passing rate of 35 only. The low achievers performed worst on two types of items—“Word Inference from Context” (Mean=17) and “Recognizing and Interpreting Details” (Mean=17). Both high achievers and low achievers performed the worst on global items this year. Both high achievers and low achievers performed best on global items measuring “Recognizing Functional Value,” with a mean of 86 for the high achievers and a mean of 48 for the low achievers. For the discrepancy between the two groups of examinees, items on “Recognizing and Interpreting Details” (D=51.67) had the highest discrimination indexes, followed by items on “Word Inference from Context” (D=42), items on “Recognizing Functional Value” (D=38), items on “Recognizing Implications and Making Inferences” (D=32.5), and items on “Recognizing Presuppositions Underlying the Text” (D=12), which had the lowest discrimination indexes.

Finally, as shown in Table 17, in the 2007 DRET, both high achievers and low achievers performed best on items measuring “Recognizing and Interpreting Details,” with an average passing rate of 73.13 for high achievers and an average passing rate of 25.75 for low achievers. Both high achievers and low achievers performed worst on items measuring “Recognizing Presuppositions Underlying the Text” (Mean=45 vs. Mean=17) for high achievers and low achievers respectively). Notice that the average passing rates of low achievers on all items this year were lower than the minimum

desirable discrimination indexes suggested by Jeng et al (1999).

In general, the results in Table 17 revealed similar findings to those of the SAET. Most of the item types reached the minimum reasonable discrimination index of 30 except for items on “Recognizing Implications and Making Inferences” (D=22) in 2002 and 2004, and items on “Recognizing Presuppositions Underlying the Text” (D=12) in 2006 and (D=28) in 2007. This suggested that most of the item types reached a satisfactory discriminatory power of at least 30 but these two item types may be somewhat ineffective in discriminating high achievers from low achievers. As a result, it was difficult to determine which type of question had the highest or lowest discriminatory power from 2002 to 2007.

Table 18. The ANOVA Analysis of Discrimination Indexes in the DRET
Dependent Variable: Discrimination Indexes

Category	Source	Type III Sum of Squares	df	Mean Square	F	Sig.
DRET	Corrected Model	6688.186(a)	28	238.864	1.223	.269
	Intercept	70487.122	1	70487.122	360.818	.000
	Item Type	3464.877	7	494.982	2.534	.028
	Year	1061.820	5	212.364	1.087	.381
	Item Type * Year	1523.713	16	95.232	.487	.941
	Error	8790.908	45	195.354		
	Total	184347.000	74			
	Corrected Total	15479.095	73			

a R Squared = .432 (Adjusted R Squared = .079)

Table 18 presents the results of the ANOVA analysis of item types, the year, and the interaction of item types and year on the discrimination indexes. The fixed factors are the year and the item types. The dependent variable is the discrimination indexes. The results of the ANOVA showed that there was significant effect on the reading skills on the discrimination index ($F=2.534, p<.05$), which indicated that the

discrimination indexes varied according to the types of reading skills measured. However, no significant effect can be found on the factor of year alone ($F=2.534$, $p<.05$) or on the interaction of both factors ($F=1.087$, $p>.05$). That is to say, there were no statistical differences among the discrimination indexes of different types of reading skill each year. In sum, although a ranking difference among the discrimination indexes on various item types existed, it was hard to determine which types of items had the most and least discriminatory power in both SAET and DRET.

Summary

In this chapter, both qualitative and quantitative analyses of the test items were presented. The qualitative analyses showed what reading skills were measured in the SAET and DRET whereas the quantitative analyses illustrated the statistical results of the passing rates (i.e., all examinees, high achievers, and low achievers) and the discrimination index on each item.

A preliminary analysis of the 2002 SAET reading comprehension test items was done to help the three raters familiarize themselves with the coding procedure and to see what skills were measured on the test. Results are reported in frequencies, descriptive statistics, as well as qualitative descriptions.

The findings of the formal analysis revealed that in the SAET and DRET from 2002 to 2007, eight types of reading skills were identified (1) Word Inference from Context, (2) Recognizing Cohesive Devices, (3) Recognizing and Interpreting Details, (4) Recognizing Functional Value, (5) Recognizing Text Organization (only in DRET), (6) Recognizing Presuppositions Underlying the Text, (7) Recognizing Implications and Making Inferences, and (8) Recognizing and Understanding the Main Idea. Throughout the six years, items on “Recognizing and Interpreting Details” were the most tested item type in both tests (61.30% on the SAET 55.40% on the

DRET). On both tests, items on “Recognizing Test Organization” were the least tested (0% on the SAET and 1.4% on the DRET). When the results of both tests are compared, the findings revealed that different types of skills tested in the SAET and DRET throughout the six years. Different reading skills appeared in both exams in different years.

Regarding the reading level which best discriminated the high achievers and low achievers, in 2002 to 2007 SAET, 5 years out of 6 were global level reading skills; for example, “Recognizing Presuppositions Underlying the Text” in 2003 (D=69), “Word Inference from Context” in 2004 (D=64) and 2005 (D=58), “Recognizing and Understanding the Main Idea” in 2006 (D=59) and 2007 (D=64). Only the reading skill that best discriminated the high achievers and low achievers in 2002 SAET was a local skill “Recognizing Cohesive Devices” (D=66). However, in the DRET, most of the skills that best discriminated the high achievers and low achievers are local skills; for example, “Recognizing Cohesive Devices” in 2003 (D=64), “Recognizing and Interpreting Details” in 2005, 2006, and 2007. In 2002 DRET, the skill that best discriminated the high achievers and low achievers was a global skill “Recognizing Text Organization” (D=72). In 2004 DRET, one local skill and one global skill both best discriminated the high achievers and low achievers, i.e., “Recognizing Cohesive Devices” (D=48.5) and “Recognizing Functional Value” (D=48.5).

In general, the overall performances of all examinees differed on various types of reading comprehension questions in the SAET and DRET. In the SAET administered from 2002 to 2007, no general pattern of the discrimination indexes of the average passing rates on each item type was found. The pattern varies each year. However, it was worth noting that items on “Recognizing Implications and Making Inferences” and items on “Recognizing and Understanding the Main Idea” had the lowest

discriminatory power in the SAET for two years. As to the DRET, items on “Recognizing Implications and Making Inferences” had the lowest discriminatory power in the DRET for 3 years (Mean=22 in 2002, Mean=43 in 2003, and Mean=22 in 2004) and items on “Recognizing Presuppositions Underlying the Text” has the lowest discriminatory power for 2 years (Mean=12 in 2006 and Mean=28 in 2007). In terms of the discrepancy between the high and low achievers, although a ranking difference among the discrimination indexes on different item types could be found, it was hard to determine which types of items had the best and worst discriminatory power in the SAET and DRET. Further discussions of these findings will be presented in detail in Chapter Five.

CHAPTER FIVE

DISCUSSION AND CONCLUSION

This chapter comprises four main sections. The first section summarizes the major findings of the present study and presents detailed discussion of these findings with reference to the results of previous findings and relevant theories. The second section provides pedagogical implications derived from the findings of the present study. Finally, limitations of the current study and suggestions for future research are provided.

The purpose of this study is to investigate the reading skills measured in the reading comprehension tests in both SAET and DRET. The reading comprehension questions in both SAET and DRET were analyzed by adopting a revised Nuttall's taxonomy. My research was guided by the following four research questions:

1. What are the reading skills measured in the SAET and DRET in the reading comprehension section in the past 6 years (from 2002~2007)?
2. What are the similarities and differences between the SAET and DRET with regard to the reading skills measured?
3. How did high school students perform on test items that measure different skills?
4. How did high achievers and low achievers differ on test items that measure different skills?

Discussion of Major Findings

The following section discusses and concludes the major findings of this study, which concern the types of reading skills measured in the SAET and DRET, similarities and differences of different types of skills measured in both exams, the examinees' performance on various types of skills, and comparisons of high and low achievers' performances on different item types.

Reading Skills Measured in the SAET & DRET

A total of 167 reading comprehension test items were analyzed by using the revised Nuttall's taxonomy as the coding scheme.

In terms of skill types, the findings revealed that seven types of reading skills were tested in the 2002 to 2007 SAET: "Word Inference from Context," "Recognizing Cohesive Devices," "Recognizing and Interpreting Details," "Recognizing Functional Value," "Recognizing Presuppositions Underlying the Text," "Recognizing Implications and Making Inferences," and "Recognizing and Understanding the Main Idea." Items that measured the examinees' abilities to recognizing text organization were not found in 2002-2007 SAET. As to the DRET, the findings showed that all the eight types of reading skills emerged: "Word Inference from Context," "Recognizing Cohesive Devices," "Recognizing and Interpreting Details," "Recognizing Functional Value," "Recognizing Text Organization," "Recognizing Presuppositions Underlying the Text," "Recognizing Implications and Making Inferences," and "Recognizing and Understanding the Main Idea." All categories of reading skills were identified on both reading comprehension tests except for items on "Recognizing Text Organization," of which only one item was found in 2002 DRET.

One thing worth noting is that "Recognizing Text Organization" is the least tested skill, with no items in the SAET and only 1 item in the DRET. A possible explanation for this might be that the articles in the reading comprehension tests are passages or excerpts from a larger context and this makes it hard for test writers to use them to write items to measure "Recognizing Text Organization."

Similarities and Differences between the Reading Skills Tested in the SAET and DRET

The similarities and differences between the SAET and DRET lay in the

frequency, occurrences, and distribution of reading skill item types. In terms of frequency, the results showed that in the SAET, 73.1% of test items measured local reading skills while only 26.9% were devoted to testing global reading skills. As for the DRET, 67.6% aimed at measuring bottom-up skills while only 32.4% were dedicated to global skills. Both exams shared a similar pattern: both exams measured more local skills. If we look at the goals of the SAT and the DRT as previously mentioned, it is reasonable to expect more global skills in the DRET. However, the results showed that both exams had around 70% of items on testing local skills and thus the results are not in accord with the testing objectives that the CEEC set for the exams

In addition, the results showed that both examinations revealed a similar pattern: the most frequently tested items were those on “Recognizing and Interpreting Details” (accounted for 61.3% on the SAET and 55.4% on the DRET), with much fewer numbers of items devoted to measuring other skills. This finding is in accord with studies on reading comprehension item analyses of the SAET and DRET (e.g., Hsu, 2005; Lan, 2007; Lu, 2002). All of the previous studies showed a consistent finding that items which aimed at measuring the examinees’ abilities to recognize and interpret detailed information are the major types. This indicates that most of the question types were lower order questions whereas higher order question types were less emphasized. A possible explanation for this might be that it is more difficult to write plausible distractors for MCQs when higher order skills are tested.

In the SAET, two types of items occurred every year, including local items on “Word Inference from Context” and “Recognizing and Interpreting Details.” By examining the SAET and DRET, it was found that items on “Recognizing and Interpreting Details” were the majority each year, ranging from 53.3% to 73.3% in the SAET and 45.5% to 72.7% in the DRET. Each year, around half of the test items were

devoted to measuring the examinee's abilities to locate specific details. Thus, it seems reasonable to conclude that both exams emphasized the development of this skill in reading. However, the concentration on assessing the more local understanding in the SAET and the DRET could well lead EFL teachers and learners to believe that lower-order reading skills are more important than the higher-order ones. From what was found in the present study, the reading skills measured in the SAET and DRET are more text-based rather than reader-based. The reason why local reading skills were over emphasized in both SAET and DRET may need further exploration in future studies.

Examinees' Performances on Various Item Types

To find out how the examinees generally performed on reading comprehension questions of various reading skills in the SAET and DRET, a two-way ANOVA were run to examine the results. The findings revealed that the ANOVA analysis of the SAET and DRET were similar. The ANOVA tests for the SAET and DRET did not found a significant effect of item types on the examinees' average passing rates.

In the SAET, "Recognizing Cohesive Device" had the highest passing rates (see Table 21). However, only one item was categorized as this type. Thus, it was difficult to see whether there is a pattern for item difficulty of this item. All the items in the SAET received high passing rates with an average of around 56.64%. Nevertheless, it was difficult to see which types of questions were best or worst performed throughout the six years. The examinees' performance varied on different types of skills each year, which indicated that the difficulty level of each item type was inconsistent every year. For example, as noted before, items on "Recognizing and Understanding the Main Idea" had the highest passing rates in 2002, 2003, and 2007 SAET while they were the lowest in 2004 SAET. Another distinct example that revealed the unrelatedness of item type and difficulty level was found in items on "Recognizing

Implications and Making Inferences.” As shown in Table 11 in Chapter Four, Items of this type ranked the lowest in 2002 (P=45), 2003 (P=41), 2007 (P=41) SAET while they were best performed in 2005 (P=66) and 2006 (P=76)SAET.

Similar to the ANOVA results for the SAET, the ranking of passing rates of different item types in the DRET were not consistent throughout the years. The findings showed that no significant effect was found in the factor of year ($F=.381$, $p>0.05$) and there was no there was no interaction between the item types and the year ($F=.941$, $p>0.05$). However, the results showed that a significant effect was found in the factor of item types ($F=2.534$, $p<0.05$), which indicated that the passing rates varied according to the types of reading skills measured on each item. However, if we examine the passing rates of different item types throughout 2002 to 2007 DRET, the passing rates of item types were not consistent throughout the years. For example, “Word Inference from Context” had a passing rate of 54 in 2002, 48 in 2003, 50.5 in 2005, but dropped to only 31 in 2006 DRET. Besides, not all types of skills were measured every year, making it impossible to determine which types of skills were the most or the least difficult ones. Moreover, different kinds of skills appeared in different years, and thus the passing rates of the skills cannot be compared.

It is noteworthy that the average passing rates of items in the SAET are higher than those in the DRET, except that the passing rates of items on “Recognizing Functional Value” in the DRET were slightly higher than those in the SAET. This difference of passing rates between the SAET and the DRET might be that the DRET is aimed to distinguish more proficient students (CEEC); thus, the DRET is expected to be more difficult than the SAET.

As Matthews (1990) argued, it would be easier for readers to reach global understanding than the local one since more redundant information was available for readers to understand the gist. The results of the passing rates on “Recognizing and

Understanding the Main Idea” in the DRET seem to be compatible with Matthew’s argument since the test takers performed much better on “Recognizing and Understanding the Main Idea” items than “Recognizing and Interpreting Details. “This could imply that the examinees in general are better at understanding the text as a whole; nevertheless, they had difficulties understanding specific detailed information.

Comparisons of High and Low Achievers’ Performances on Different Item Types

Regarding the reading skill which best discriminated the high achievers and low achievers, in 2002 to 2007 SAET, three local reading skills best discriminated high and low achievers and three global reading skills best discriminated high and low achievers. In 2002, 2004, and 2005, local reading skills had the highest discrimination indexes; for example, “Recognizing Cohesive Devices” (D=66) in 2002 SAET, “Word Inference from Context” in 2004 (D=64), and “Word Inference from Context” in 2005 SAET. This indicated that local items were easy for high achievers but difficult for low achievers in these years. In contrast, in 2003, 2006, and 2007, global reading skills had the highest discrimination indexes; for example, “Recognizing Presuppositions Underlying the Text” (D=69) in 2003, “Recognizing and Understanding the Main Idea” (D=59), and “Recognizing Implications and Making Differences” (D=64). This indicated that global items were easy for high achievers and difficult for low achievers.

Similar to the results in the SAET, most of the skills that best discriminated the high achievers and low achievers in the DRET are local skills as well; for example, “Recognizing Cohesive Devices” in 2003 (D=64), “Recognizing and Interpreting Details” in 2005, 2006, and 2007. However, In 2002 DRET, the skill that best discriminated the high achievers and low achievers was a global skill “Recognizing Text Organization” (D=72). In 2004 DRET, one local skill and one global skill both

best discriminated the high achievers and low achievers, i.e., “Recognizing Cohesive Devices” (D=48.5) and “Recognizing Functional Value” (D=48.5). In sum, in both SAET and DRET, most of the skills that best discriminated high and low achievers were local skills. This suggested that both in the SAET and DRET, questions testing local skills were easy items for both the high achievers but difficult for the low achievers, and items on global skills were difficult for both high and low achievers.

In general, in the SAET from 2002 to 2007, the discrimination indexes of all item types reached the ideal discrimination index of 30 established by Jeng (1999) and most of the discrimination indexes were far above the minimum desirable index. This indicated that all of the items in the SAET seemed to be easy for high achievers but difficult for low achievers, meaning all of the items truly discriminated high achievers and low achievers. Also, most of the items had rather good discriminatory power since they had discrimination indexes far above the ideal index. However, the findings in the DRET from 2002 to 2007 showed a different pattern. In the DRET, the discrimination indexes of most items were above the ideal discrimination index of 30 but some items failed to reach the minimum desirable index, such as global items on “Recognizing Implications and Making Inferences” (D=22 in 2002, and D=22 in 2004) and “Recognizing Presuppositions Underlying the Text” (D=12 in 2006 and D=28 in 2007). As aforementioned, most of the skills that best discriminated high and low achievers in the DRET were local skills. Thus, the results indicated that these two types of items were probably too difficult for most examinees and did not appropriately distinguish the high achievers from low achievers. This finding indicated that there is a need for the CEEC to be more careful when designing items to assess students’ abilities to make inferences and understand the presuppositions underlying the texts.

Conclusions

In this section, we will first present a summary of the major findings. Then, pedagogical implications based on the results and limitations and suggestions for future research will be provided.

Summary of the Major Findings

The present study was conducted to explore the reading skills measured on the reading comprehension test items of the SAET and DRET administered from 2002 to 2007. Furthermore, the examinees' performances on each skill type were analyzed to see their strengths and weaknesses while taking reading comprehension tests. The main findings related to the research questions were summarized as follows.

First, the major findings of the study revealed that a total of eight types of reading skills were identified in the SAET and DRET from 2002 to 2007. These skills were "Word Inference from Context," "Recognizing Cohesive Device," "Recognizing and Interpreting Details," "Recognizing Functional Value," "Recognizing Text Organization," "Recognizing Presuppositions Underlying the Text," "Recognizing Implications and Making Inferences," "Recognizing and Understanding the Main Idea. Eight types of reading skills were identified on both reading comprehension tests except for items on "Recognizing Text Organization." Only one item was labeled as "Recognizing Text Organization" in 2002 DRET.

Second, items on "Recognizing and Interpreting Details," a more local skill, were the major types in the SAET and DRET throughout the six years. This indicated that local reading skills instead of global reading skills were favored in both exams.

Third, the major differences lay in the frequency, occurrence, and distribution of items measuring different types of reading skills. In the SAET, two types of skills appeared each year: items on "Word Inference from Context," and items on "Recognizing and Interpreting Details." As to the DRET, only item on "Recognizing

and Interpreting Details” were tested every year. This indicated that local reading skills instead of global ones were favored in both exams.

Finally, in both exams, most item types reached the ideal discrimination index of 30 established by Jeng (1999) and most even had discrimination index far above the minimum desirable discrimination index. However, two item types in the DRET failed to reach the minimum desirable index, such as items on “Recognizing Implications and Making Inferences” (D=22 in 2002, and D=22 in 2004) and items on “Recognizing Presuppositions Underlying the Text” (D=12 in 2006 and D=28 in 2007).

Pedagogical Implications

The findings of the current research, which explores the reading skills measured in the SAET and DRET, had some pedagogical implications for reading instruction and testing in senior high schools. One implication drawn from the results is that by knowing what reading skills are measured on these two exams could help teachers to know what skills are required when taking the SAET and DRET and to enhance their students’ reading ability by teaching them how to use these eight reading skills: (1) “Word Inference from Context,” (2) “Recognizing Cohesive Devices,” (3) “Recognizing and Interpreting Details,” (4) “Recognizing Functional Value,” (5) “Recognizing Text Organization,” (6) “Recognizing Presuppositions Underlying the Text,” (7) “Recognizing Implications and Making Inferences,” (8) “Recognizing and Understanding the Main Idea.” In addition, by knowing what item types the examinees performed poorly, teachers could help their students practice these skills more.

In the DRET, the examinees performed the worst on “Recognizing Presuppositions Underlying the Text,” with the passing rates lower than the standard ones set by Jeng (1999). Teachers then are highly recommended to improve students’

abilities of recognizing a writer's purpose, attitude, mood, and tone while reading texts. Teachers need to help the students identify related details, main ideas, and cause-effect relationships in order to make appropriate inferences and generalizations (Dillner & Olson, 1982). For example, to raise the students' awareness, as Dillner & Olson (1982) suggested, teachers may have them compare two articles written by different writers on the same topic in order to distinguish each writer's purpose. As Nuttall (1996) stated, presupposition is bound up with inference. One thing important to be noted is that "Recognizing Implications and Making Inferences" also obtained low passing rates in the DRET. Making inferences requires readers to use their knowledge and it is often regarded as an advanced skill and tends to be overlooked (Nuttall, 1996). Teachers are strongly recommended to train the students to select and relate relevant facts stated in texts in order to synthesize unstated meaning in texts. Teachers can also use questions of inference to train students to consider what is implied but not explicitly stated in texts (Nuttall, 1996). Moreover, teachers need to encourage their students to make use of syntactic, logic, and cultural clues to understand the meaning of unknown elements in texts (Grellet, 1982).

Additionally, five of the item types were much worse performed on the DRET than on the SAET, with a gap of discrimination index ranging from 10 to 27. These skills were "Recognizing Cohesive Devices," "Recognizing and Interpreting Details," "Recognizing Presuppositions Underlying the Text," "Recognizing Implications and Making Inferences," and "Recognizing and Understanding the Main Idea." Of these five skills, "Recognizing Cohesive Devices" had the largest gap (27). In a study of college freshmen's comprehension and application of text cohesion, Huang (1993) argued that most college students were poor at detecting cohesion in English texts. Thus, to better prepare students for the DRET, teachers should help them recognize cohesive devices and practice this skill while reading.

Another important implication derived from the current research is that “Recognizing Text Organization” should be tested more often since only one item was labeled as this type in the DRET and none was found in the SAET. As stipulated in the curriculum guidelines for senior high school English instruction, students are required to be able to understand the organization of texts. If the SAET and DRET do not test what teachers are trying to teach, it’s very likely that students will not pay attention to what teachers teach and eventually teachers will stop teaching it (Nuttall, 1996). As a result, the present study calls for more attention for the CEEC to include more items that measure students’ abilities to recognize and understand textual organization in the SAET and DRET.

In addition, as revealed in the findings, in both SAET and DRET, items on local skills best discriminated high and low achievers. Global items seemed too difficult for all examinees and a few global items even had discrimination index lower than the ideal index suggested established by Jeng (1999). Hence, it is suggested that teachers should teach students global skills and to help them practice those skills. Lastly, as the findings revealed, both SAET and DRET favored items on local skills. This could lead teachers and students to believe that local items are more important than global items. Thus, it is suggested that test writers write more items which measure global skills in the SAET and DRET so that a washback would be formed to urge teachers to familiarize students with global skills.

Limitation and Suggestions for Future Research

This study explored the reading skills tested in the SAET and DRET. In this section, limitations of the present study are presented. First, item analysis in the present study can only help to predict the reading skills attempted to be measured on each item. No experimental measures were taken to probe into the examinees’ minds to investigate their reading process and to see whether the examinees in fact apply the

reading skills when answering a question. To do that would require a think-aloud method. For future research, it is suggested that researchers or teachers adopt a think-aloud method to find out how the students arrive at a particular answer and why they think the answer is correct. By doing this, we may know whether the item intended to test a certain type of skill truly measures that skill.

Second, whether teachers teach these reading skills at school would also be worthy of future research. Through this, we could know whether the exams are in accord with the real teaching situations in the classrooms and to see whether reading skill instruction truly help students.

Third, the SAET and DRET had different samplings. There were more examinees taking the SAET than the DRET. As previously mentioned in Chapter Three, the SAET might have had screened some proficient students and those students may choose not to take the DRET. This might have resulted in the different performances of passing rates on various items. However, discussion of the possible influence of different sampling on the passing rates is beyond the scope of this paper.

Lastly, the current study aimed to categorize reading comprehension test items in the SAET and DRET. It is suggested that future research adopt the taxonomy in the present study to analyze the reading comprehension questions in senior high school English textbooks to see whether they were written in order to help students practice reading skills. Finally, it is also suggested that future research attempt to analyze the curricula for senior high school English reading to see whether there is an alignment of curriculum, teaching, and assessment of reading.

REFERENCES

- Abbott, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. *Language Testing*, 24(1), 7-36
- Alderson, J. C. (2000). *Assessing reading*. UK: Cambridge University Press.
- Anderson, N. J., (1999). *Exploring second language reading: Issues and strategies*. Boston, MA: Heinle & Heinle Publishers.
- 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.
- Bernhardt, E.B. (1991). *Reading development in a second language: Theoretical, empirical and classroom perspectives*. Norwood, Ablex Publishing Corporation, N.J.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. 2nd ed. New York: Longman.
- Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. New York: Longman.
- Carrell, P. L. (1983a). Some issues in studying the role of schemata, or background knowledge in second language comprehension. *Reading in a Foreign Language*, 1(2), 81-92.
- Carrell, P. L. (1988a). Introduction: Interactive approaches to second language reading. In Carrell, P. L., Devine, J. & Esdey, D. (Eds.), *Interactive approaches to second language reading* (pp.1-7). Cambridge: Cambridge University Press.
- Carrell, P. L. (1988b). Some causes of text-boundedness and schema interference in ESL reading. In Carrell, P. L., Devine, J. & Eskey, D. (Eds.), *Interactive approaches to second language reading* (pp.101-113). Cambridge: Cambridge

University Press.

- Carrell, P. L., & Eisterhold, J. (1983). Schema theory and ESL reading pedagogy. In M.H. Long and J.C. Richards (Eds.), *Methodology in TESOL: A book of reading* (pp. 218-232). Boston, Mass: Heinle & Heinle Publishers.
- Carver, D. (1978). Reading comprehension—Is there such a thing,” *English Language Teaching*, 32(4), 291-297
- Chen, K. T. (1995) *英文科試題分析 [An item analysis of the 1995 SAET]*. Taipei: CEEC., 38-49
- Chen, J. C (2005). *Explicit instruction of reading strategies at senior high school in Taiwan*. Unpublished Master Thesis. Kaohsiung: National Kaohsiung Normal University.
- Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Hackay, B. Brakeman, and R.R. Jordan (Eds.), *Reading in a second language* (pp. 5-12). Rowley, MA: Newbury House Publishers.
- Cohen, A. D. (1994). *Assessing language ability in the classroom* (2nd ed.). Boston: Newbury House/Heinle & Heinle.
- Cohen, A.D. (1998). Strategies and process in test taking and SLA. In L.F. Bachman and A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 90-111). Cambridge: Cambridge University Press.
- Cummings, O. W. (1982). Differential measurement of reading comprehension skills for students with discrepant subskill profiles. *Journal of Education Measurement*, 19(1), 59-66
- Dillner, M. H. & Olson, J. P. (1982). *Personalizing reading instruction in middle, junior, and senior high schools: Utilizing a competency-based instructional system*. New York: Macmillan.

- Dubin, F.D., Eskey, D. E. & Grabe, W. (1986). *Teaching second language reading for academic purposes*. Reading, Mass: Addison-Wesley.
- Dunmore, D. (1989). Using Contextual Clues to Infer Word Meaning: An Evaluation of Current Exercise Types. *Reading in a Foreign Language*, 6 (1), 337-347
- Eskey, D. E. (1988). Holding in the bottom: an interactive approach to the language problems of second language readers. In Carrell, P. L., Devine, J. and Esdey, D. (Eds.), *Interactive approaches to second language reading* (pp. 93-100). Cambridge: Cambridge University Press.
- Goodman, K.S. (1967). Reading: A psycholinguistic guessing game. In H. Singer & R. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, Delaware: International Reading Association.
- Gough, P. B. (1972). *One second of reading*. In Kavenagh, F.J. and Mattingly, G. (Eds.), pp. 331-358.
- Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25(3), 375-406.
- Grabe, W. (1986). The transition from theory to practice in teaching reading. In F. Dubin, D. E. Eskey, & W. Grabe (Eds.), *Teaching second language reading for academic purposes* (pp. 25-48). Reading, MA: Addison-Wesley Publishing Company.
- Grellet, F. (1981). *Developing reading skills: A practical guide to reading comprehension exercises*. Cambridge: Cambridge University Press.
- Farr, R., Prichard, 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
- Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw Hill
- Harrison, A. (1983). *A language testing handbook*. London: Macmillan.

- Heaton, J. B. (1988). *Writing English language tests*. New York: Longman.
- Hsu, W. L. (2005). *An analysis of the reading comprehension questions in the JCEE English test*. Unpublished master thesis: Kaohsiung: National Kaohsiung Normal University.
- Huang, T. L. (1993). 大學生英文選詞連句成段能力之研究 [An analysis of college students' abilities to use cohesive ties], *New Horizons in English Language Teaching*. (pp. 202-241). Taipei: The Crane Publishing Company.
- Huang, T.S. (1994). *A qualitative analysis of the JCEE English tests*. Taipei: The Crane Publishing Company.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Jeng, H. S. (1992). 中華民國八十年度大學年考英文選擇題統計數值分析 [A Statistical Analysis of English Test Items in 1991 JCEE], presented at the First International Symposium on English Teaching in ROC., held in Taipei, included in the Proceedings, 251-271.
- Jiang, W. J., & Lin, B. F. (1999). 八十八學年度大學聯考英文試題簡析[A brief analysis of English test items in 1999 JCEE]. *English Teaching and Learning*, 24(1), 3-23
- Kirby, J. R. (1988). Style, strategy, and skill in reading. In R.R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 229-274). New York: Plenum Press.
- Lan, W. H. (2007). *An analysis of reading comprehension questions on the SAET and the DRET using revised Bloom's Taxonomy*. Unpublished master's thesis. Taipei: National Taiwan Normal University.
- Lu, J. Y. (2002). *An analysis of the reading comprehension test given in the English Subject Ability Test in Taiwan and its pedagogical implications*. Unpublished

- master's thesis. Taipei: National Chengchi University.
- Mackay, R. (1987) Teaching the information gathering skills. In M.H. Long and J.C. Richards (Eds.), *Methodology in TESOL: A book of reading* (pp. 248-258). Boston, Mass: Heinle & Heinle Publishers.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Madsen, H. S. (1983) *Techniques in Testing*. New York: Oxford University Press.
- Munby, J. (1978) *Communicative Syllabus Design*. Cambridge, Cambridge University Press.
- Mo, C. C. (1987). A study of English reading comprehension and general guidelines for testing reading. *Journal of National Chengchi University*, 55, 173-206. Taipei: National Chengchi University.
- Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, 6, 199-215.
- Nuttall, C. (1996). *Teaching reading skills in a foreign language*. London: Heinemann.
- Omaggio, A. H. (2001). *Teaching language in context*. Boston: Heinle & Heinle.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20, 26–56.
- Pressley, M. (2000). Comprehension Instruction: What makes sense now, what might make sense soon. *Handbook of Reading Research*. Vol. 3. Retrieved February 18, 2008, from <http://www.readingonline.org/articles/handbook/pressley/>
- Purpura, J. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 47, 289–325.
- Raygor, A. L. & Raygor , R. D. (1985). *Effective reading: Improving reading rates*

and comprehension. New York: McGraw-Hill.

Richards, J. C. (1997). *From reader to reading teaching: Issues & strategies for second language classrooms*. Cambridge: Cambridge University Press.

Richards, J. C. & Schmidt, R. (2002). *Longman dictionary of language teaching & applied linguistics* (3rd ed.). Pearson Education Limited.

Richards, J. C. (2003). *30 years of TEFL/TESL: A personal reflection*. Retrieved September 17, 2007, from

www.professorjackrichards.com/pdfs/30-years-of-TEFL.pdf

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic, (Ed.), (pp. 573-603). *Attention and Performance VI*. Hillsdale, NJ.: Erlbaum.

Rumelhart, D. E. (2004). Toward an interactive model of reading. In R. B. Ruddell and N. J. Unrau (5th edition), *Theoretical models and processes of reading*. Newark, Delaware: International Reading Association.

Sequera, W. (1995). Construct Validity in Reading Tests. *Forum*, 33(1), 49

Shi, Y. H. (1991). Teaching to the Test: A Case Study of Taiwan. *Studies in English Literature & Linguistics*, 1991, 51-63

Slomp, H. D. (2005). Teaching and assessing language skills: Defining the knowledge that matters. *English Teaching: Practice and Critique*, 4(3), 141-155

Smith, F. (1971). *Understanding reading: A psycholinguistic analysis of reading and learning to read*. New York: Holt, Rinehart & Winston.

Stanovich, K. E. (1980). Toward an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16(1), 32-71.

Su, C. (2001). Evaluation of reading skills applied in the first-year college English Reading Course. Selected Papers from the 10th International Symposium on English Teaching. (pp. 579-587). Taipei: The Crane Publishing Co.

- Teale, W. H. & Rowley, G. (1984). Standardized Testing and the teaching of reading: A practical guide with evaluations of reading tests commonly used in Australian schools. San Antonio: The University of Texas, Division of Education School of Social and Behavioral Sciences. (ED 291 769).
- Urquhard, A. H. & Weir, C. J. (1998). *Reading in a second language: Process, product and practice*. London: Longman.
- Wallace, C. (1992). *Reading Oxford*, Oxford University Press.
- Weir, C. J. (1990). *Communicative language testing*. New Jersey: Prentice Hall.
- Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.
- Williams, R. (1983). Teaching the Recognition of Cohesive Ties in Reading a Foreign Language. *Reading in a Foreign Language, 1*(1), pp35-52
- 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.
- Williams, R. (1983). Teaching the recognition of cohesive ties in reading a foreign language. *Reading in a Foreign Language, 1*(1), 35-52
- Xu, Y. J., & Lu, S. F. (1998). 八十七學年度大學聯考英文試題簡評[A brief criticism of English test items in 1998 JCEE]. *English Teaching and Learning, 23*(2), 23-40

Appendix 1 — Principles for Classification of the Reading Skills Based on the Revised Nuttall’s Taxonomy

Principles for classification of reading skills based on the revised Nuttall’s taxonomy are provided as follows and example questions were selected to illustrate how the skills are measured. To categorize questions in the reading passages, the stem of each question, the correct answer, and the required information in the reading passages to answer the question are processed in order to classify the items into a certain type of ESL/EFL reading skills. The items were thus coded according to the reading skill that the raters believed to be instrumental at arriving at the answer.

1. *Word Inference from Context*

An item is categorized as *Word Inference from Context* when it requires test takers to determine unfamiliar words or phrases by using contextual clues which proceed or follow them from context (See Example 1).

Example 1

What does “**the world**” mean in “show the world to the world”?

- *(A) The global village.
- (B) The leading museums.
- (C) The ancient civilizations.
- (D) The international public.

(Q 47 on the 2004 DRET)

2. *Recognizing and Interpreting Cohesive Devices*

An item is categorized as *Recognizing and Interpreting Cohesive Devices* when it requires test takers to recognize and interpret cohesive devices. Items that measure the readers’ ability to interpret pro-forms, elliptical expressions or lexical cohesion are considered to measure this skill. To identify pro-forms, readers are required to identify the meanings of words like *it*, *our*, *this*, *those*, *then*, *one* (as in *the wrong one*), *so/no* (as in *I think so, it appears not*), and comparatives (*smaller*,

same, additional, such, other, etc.). To interpret elliptical expressions, test takers are required to retrieve the meaning of something left unsaid but has already been mentioned in the text previously. To interpret lexical cohesion, test takers are required to interpret the relationship between a lexical item and other parts of the discourse (e.g. synonyms, hyponyms, metaphors, etc.). Items that measure the readers' ability to interpret pro-forms, elliptical expressions or lexical cohesion are considered to measure this skill. See Example 2 for a sample item of this type of questions.

Example 2

Who does *he* in line 2, paragraph 3 refer to?

- (A) The policeman.
- (B) Soapy.
- * (C) The manager.
- (D) The judge.

(Q 56 on the 2002 DRET)

3. *Identifying and Interpreting Details*

An item is categorized as *Identifying and Interpreting Details* when test takers are required to locate and interpret details (e.g. facts, information, statements, etc.) explicitly stated in the text. For example, to answer Example 3 as follows, test takers need to understand the question and its four options and then match the options with the details specified in the text. The information of the option which cannot be found in the text would be the answer to this question.

Example 3

Which of the following is NOT a reason for the Japanese to use more packaging?

- (A) Packaging helps to keep food fresh.
- (B) Packaging helps the Japanese to show appreciation
- (C) The Japanese consider packaging a symbol of quality.
- * (D) Packaging is a way to compete with Westerners in economy

(Q 47 on the 2006 DRET)

4. *Recognizing Functional Value*

An item is categorized as *Recognizing Functional Value* when it asks examinees to recognize and understand the functional value of a sentence or a passage. Three categories of functional value can be identified: (1) independent functions (associated with propositional meaning; e.g. *defining, classifying, generalizing, naming, describing, reporting, speculating, predicting, etc.*), (2) text-dependent functions (associated with contextual meaning; e.g. *asserting, exemplifying, explaining, reinforcing, explicating, hypothesizing, commenting, concluding, etc.*), and (3) interaction-dependent functions (associated with pragmatic meaning; e.g. *conceding, evaluating, inviting, instructing, apologizing, suggesting, complaining, complimenting, warning, etc.*). See Example 4 for a sample item.

Example 4

The purpose of this passage is to

- (A) explain how people become bilingual.
- (B) explain how to be a better second language learner.
- * (C) describe research into the brains of bilingual people.
- (D) describe the best ways to acquire languages at different ages.

(Q 53 on the 2006 SAET)

5. *Recognizing Text Organization*

An item is categorized as *Recognizing Text Organization* when it requires test takers to recognize the rhetorical patterns of the passage (e.g. definition, description, cause-effect, classification, comparison-contrast, process, hypothesis, argumentation, exemplification, etc.) and analyze the rhetorical organization of a text. See Example 5 for an item coded under this category.

Example 5

This passage is written mainly to _____.

- * (A) compare
- (B) tell a sad story

- (C) amuse the readers
- (D) report a scientific discovery

(Q 53 on the 2002 DRET)

6. *Recognizing the Presuppositions Underlying the Text*

An item is categorized as *Recognizing the Presuppositions Underlying the Text* when it requires test takers to use the presented information in the passage to identify either the knowledge and experience that the writer expects the reader to have or the author's opinions, attitudes, emotions that the writer expects the reader to share, or at least to understand. The following is an example of this type of items.

Example 6

What is the author's attitude towards blood diamonds?

- (A) Indifferent.
- (B) Threatening.
- (C) Sympathetic.
- * (D) Disapproving.

(Q 56 on the 2007 DRET)

7. *Recognizing Implications and Making Inferences*

An item is categorized as *Recognizing Implications and Making Inferences* when it requires examinees to infer unstated idea, to draw unstated conclusions, or to find implications from presented information in the passage (See Example 7).

Example 7

What can we infer from William Warren's comment on Bangkok at the end of the passage?

- (A) Bangkok is a place where visitors can buy many souvenirs.
- (B) Bangkok offers attractive travel insurance to visitors.
- * (C) Bangkok is an exciting place to visit.
- (D) Bangkok reminds Warren of home.

(Q 48 on the 2003 DRET)

8. *Recognizing and Understanding Main Idea*

An item is categorized as *Understanding Main Idea* when it requires examinees to have a thorough understanding of the whole passage and to find the central idea of the text as in the following example.

Example 8

What is the main idea of this passage?

- *(A) The city of Bangkok is urbanized but it is also rich in traditional Thai culture.
- (B) Visitors to Bangkok might find the weather, the heat, and floods unbearable.
- (C) Bangkok is an international city, just like Paris, London, and New York.
- (D) There are a variety of restaurants and social events in Bangkok.

(Q 46 on the 2003 DRET)

Appendix 2—Results of the Formal Item Analysis

Instructions: Please identify the reading skills measured on each item based on its stem, the correct answer, and the information in the passage needed to answer a question. Mark the reading skills by their identification.

Reading Skills												
SAET							DRET					
#	2002	2003	2004	2005	2006	2007	2002	2003	2004	2005	2006	2007
41	1	8	8	4	4	4						
42	1	7	1	2	2	1						
43	3	2	1	7	1	1						
44	1	8	2	1	1	1						
45	7	1	1	4	1	1						
46	8	1	1	1	1	1		8	4	1	4	1
47	1	1	2	1	1	1		1	3	1	1	1
48	2	1	8	1	7	1		7	7	1	2	7
49	1	1	1	1	8	7		1	4	2	1	1
50	7	1	1	1	2	2		4	1	8	7	1
51	2	6	1	1	1	1	4	1	3	1	1	1
52	1	8	1	4	1	6	7	2	1	1	1	1
53	1	1	1	1	4	8	5	1	1	2	6	1
54	1	1	1	1	1	1	8	1	1	8	1	7
55	1	7	1	1	1	6	1	1	1	8	1	1
56				4	1	2	3	1	1	8	7	6
57							1	1				
58							1	3				
59							1	7				
60							1	4				
61							2					
62							1					
63							1					
64							1					
65							6					

Note. 1 refers to “Word Inference from Context.” 2 refers to “Recognizing Cohesive Devices.” 3 refers to “Recognizing and Interpreting Details.” 4 refers to “Recognizing Functional Value.” 5 refers to “Recognizing Text Organization.” 6 refers to Recognizing Presuppositions Underlying the Text.” 7 refers to “Recognizing Implications and Making Inferences.” 8 refers to “Recognizing and Understanding the Main Idea.”